

Mémoire de fin d'étude pour l'obtention du

Master professionnel
Statistique, Informatique et Techniques Numériques

Modélisation du développement et de la phénologie du manguier

Alexandra JESTIN
6 septembre 2013

Maître de stage : Frédéric NORMAND
Co-encadrant : Frédéric BOUDON

Résumé

L'unité HortSys du CIRAD (Centre de coopération internationale en recherche agronomique pour le développement) mène depuis plusieurs années à la Réunion des travaux sur le manguier afin de proposer des modes de conduite favorisant un rendement régulier et une phénologie synchrone, tout en limitant l'utilisation d'intrants et en améliorant la qualité des fruits. L'objectif de ce stage, est de construire un modèle de phénologie du manguier basé sur les connaissances acquises sur le développement de cet arbre.

Table des matières

Introduction	3
1 Le manguier	5
1.1 Concepts et définitions biologiques	5
1.1.1 La phénologie	5
1.1.2 Le cycle de production et le cycle phénologique	5
1.1.3 Croissance séquentielle et unité de croissance	7
1.1.4 Position apicale, latérale et terminale	7
1.1.5 Inflorescence	8
1.1.6 Nature végétative vs. nature florifère	8
1.1.7 Croissance séquentielle vs. réitération	9
1.1.8 Description schématique de la ramification	9
1.2 Matériels	11
1.2.1 Le verger et les données	11
1.2.2 Etat de l'art des analyses de la base Doralice	12
2 Le contexte des logiciels	16
3 Vers l'architecture du manguier...	18
3.1 Simuler avec le langage des L-systèmes	18
3.1.1 La structure	18
3.1.2 La géométrie : représentation 3D, la Tortue	19
3.1.3 La dynamique : évolution de la chaîne de caractères	21
3.2 Un modèle simple de croissance du manguier	25
4 Modélisation basée sur les modèles linéaires généralisés (GLM)	31
4.1 Le modèle statistique	31
4.2 Application du modèle au manguier	35
4.2.1 Les variables du modèle	35
4.2.2 Application des GLM	36
4.2.3 La prédiction du paramètre de la distribution réponse	37
4.2.4 Simulation du nombre d'UCs filles et leur date d'apparition	38
4.2.5 Modélisation de la croissance dans L-py	39
5 Modélisation basée sur les modèles graphiques	41
5.1 Le modèle statistique	41
5.2 Application du modèle au manguier	43
5.3 Les procédés de simulation	47
6 Résultats préliminaires	51
6.1 Méthode de validation préliminaire	51
6.2 Premiers résultats	53
6.2.1 Le nombre d'UCs produites au cours d'un cycle	53
6.2.2 La distribution du nombre d'UCs filles par UC mère	55
6.2.3 La distribution des dates d'apparition des UCs	57

6.2.4	La distribution du taux d'UC terminale en position apicale	59
6.2.5	La distribution du taux de floraison	61
Conclusion		63
Bibliographie		64
Annexes		66
A Symboles interprétés par la tortue		66

Introduction

Mon stage s'est déroulé en deux parties. La première partie s'est déroulée au sein de l'unité Hortsys situé à Saint-Pierre, La Réunion. Les chercheurs de cette unité s'intéressent à l'amélioration de la conduite de systèmes horticoles en milieu tropical. L'unité s'intègre dans le projet PIMAN (production intégrée de la mangue et de l'ananas) qui étudie le manguier. La deuxième partie du stage s'est déroulée au sein de l'équipe Virtual Plants à Montpellier. L'équipe se situe dans le domaine de la bio-informatique. C'est une équipe née d'un partenariat entre le CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), l'INRIA (Institut National de Recherche en Informatique et en Automatique) et l'INRA (Institut National de Recherche en Agromonie). L'équipe étudie la croissance et la structure des arbres, aussi bien à l'échelle macroscopique que microscopique. Mon stage avait pour but de modéliser la phénologie du manguier.

Cultivé depuis plus de 4000 ans et originaire du nord de l'Inde, le manguier est un arbre qui peut vivre plusieurs centaines d'années. C'est l'une des espèces fruitières tropicales les plus cultivées dans le monde et qui fait l'objet d'un commerce international en expansion. La mangue représente la cinquième production fruitière mondiale : 38,7 MT produites en 2010 (Gerbaud, 2012). Cette expansion montre à quel point cet arbre et son fruit sont appréciés.

Sur l'île de la Réunion, la mangue représente la quatrième espèce fruitière la plus plantée après l'ananas, les agrumes et le litchi. La production (environ 3 000 - 3 500 T par an) alimente principalement le commerce local et une petite quantité est exportée vers la métropole.

Cependant, le manguier présente un certain nombre de problèmes agronomiques. Parmi ceux-ci, l'alternance de production d'une année sur l'autre (une année de bonne production est suivie d'une année de faible production) et les asynchronismes phénologiques (dans le temps, la floraison par exemple est très étalée sur un manguier et entre les manguiers d'un verger) sont récurrents dans toutes les zones de production du monde. Ils sont principalement liés à une floraison "capricieuse". Ce phénomène pose problème pour deux raisons principales. Cela implique la nécessité de main-d'œuvre sur une plus longue période pour la récolte des fruits et cela nécessite des traitements phytosanitaires à plusieurs reprises au cours du cycle de production. C'est pourquoi de nombreux travaux s'accordent à mieux appréhender la physiologie du manguier et les facteurs conditionnant sa floraison et sa fructification. A la Réunion, l'objectif principal du CIRAD est d'améliorer la qualité et le rendement du fruit tout en préservant l'environnement et la biodiversité. C'est le concept de production fruitière intégrée (Normand et al., 2011).

Les travaux menés par le CIRAD depuis quelques années à la Réunion ont permis de mieux comprendre comment apparaissent et se maintiennent les asynchronismes phénologiques (Dambreville et al., 2013a). Ils résultent d'interactions entre les composantes structurelles et temporelles de l'architecture, et entre la croissance végétative et la floraison. Ces interactions peuvent aussi expliquer l'alternance de floraison, et donc de pro-

duction. Elles sont de plus affectées par des variables climatiques, notamment la température (Dambreville et al., 2013b). Ces interactions sont complexes et affectées par différents facteurs et la modélisation apparait comme un outil de choix pour synthétiser ces connaissances et explorer par la simulation le fonctionnement végétatif et reproducteur du manguiers.

L’objectif de ce stage est de synthétiser les connaissances acquises sur l’architecture du manguiers et son développement, en modélisant la croissance du manguiers avec le formalisme des L-systèmes. Pour caractériser la temporalité d’apparition de nouveaux éléments sur l’arbre, nous avons développé deux approches complémentaires. La première approche est guidée par une analyse biologique de la croissance du manguiers et utilise les modèles linéaires généralisés pour caractériser les processus de mise en place des nouveaux organes. La seconde approche est basée sur une caractérisation de la nature et de la temporalité des entités du manguiers et utilise les modèles graphiques pour modéliser leur inter-dépendance.

Nous allons dans un premier temps présenter toutes les notions nécessaires à la compréhension de la croissance végétative et reproductive du manguiers, les méthodes informatiques et les outils statistiques permettant la modélisation de l’arbre. Puis, nous présenterons un système de simulation avec les deux approches de simulation et ensuite nous exposerons les résultats obtenus. Enfin, nous discuterons des perspectives des modèles.

1 Le manguiier

1.1 Concepts et définitions biologiques

Plusieurs termes, à caractère agronomique ou propres à l'étude, seront utilisés dans ce rapport pour illustrer nos propos et nos orientations scientifiques. Pour mieux appréhender leur sens, ils sont définis dans cette partie.

1.1.1 La phénologie

La phénologie est une science ayant pour objet l'étude des phénomènes qui marquent le développement des plantes au cours de l'année : le débourrement des bourgeons, la croissance végétative, la floraison, la fructification, ...

1.1.2 Le cycle de production et le cycle phénologique

Le **cycle de production** est l'enchainement dans le temps des principales étapes de la croissance végétative et de la reproduction. Le cycle de production du manguiier dure un an et demi, il se décompose en quatre périodes importantes, également appelées stades phénologiques : la croissance végétative, le repos végétatif, la floraison et la fructification.

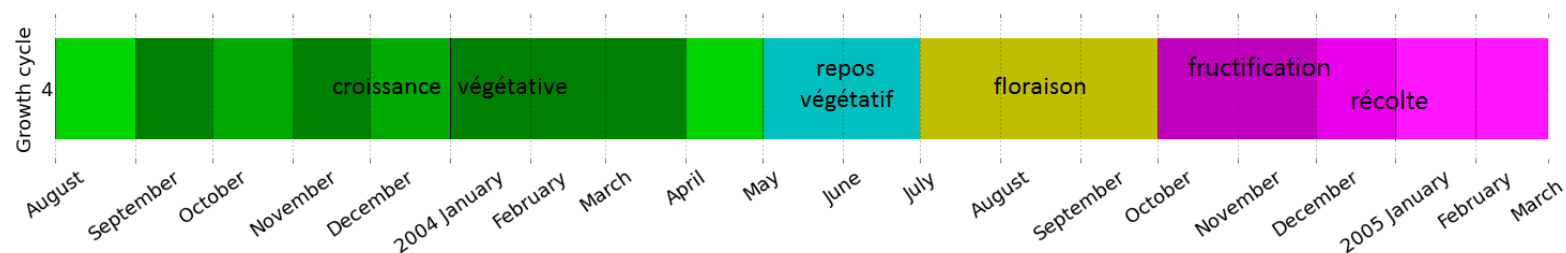


FIGURE 1 – Cycle de production du manguiier pour la variété Cogshall, à Saint-Pierre, La Réunion

La figure 1 indique le cycle de production du manguiier pour la variété Cogshall, située à Saint-Pierre, La Réunion. La nuance de la couleur verte montre l'intensité de la croissance végétative. Lorsque la couleur est claire, il y a peu de croissance. La nuance de la couleur violette indique l'intensité de la récolte. Lorsque la couleur est claire, l'arbre est moins chargé en fruits.

Le début de la croissance végétative est marqué par une faible croissance car il y a la superposition des cycles de production sur l'arbre, et une probable compétition entre la floraison et la fructification du cycle précédent, et le début de la croissance végétative du cycle suivant.

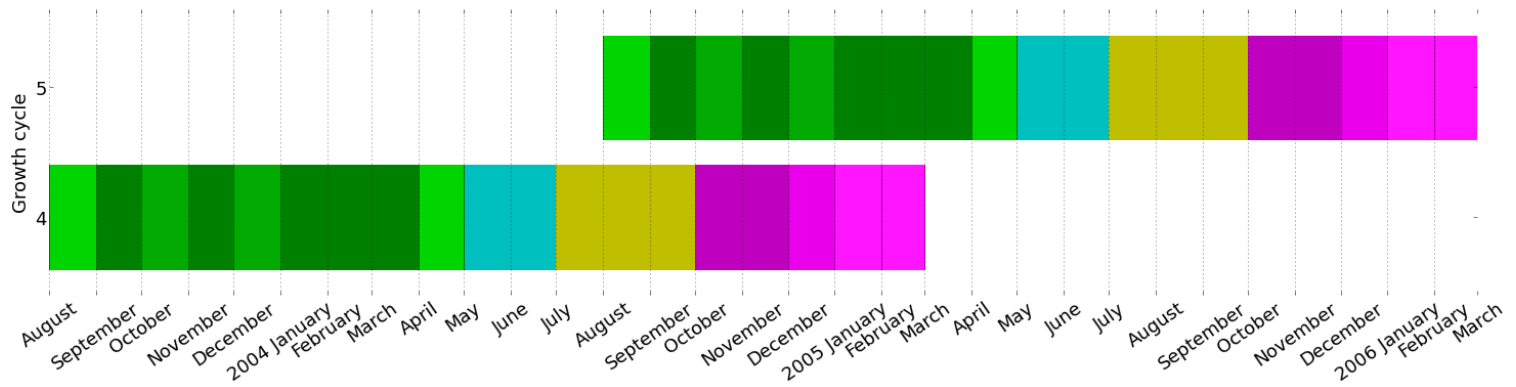


FIGURE 2 – Superposition des deux cycles de production étudiés

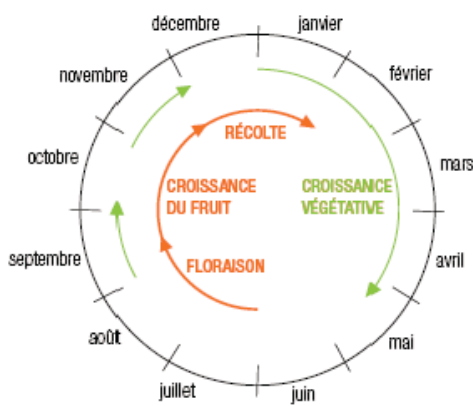


FIGURE 3 – Le cycle phénologique du manguiier, variété Cogshall, à Saint-Pierre, La Réunion (source : Vincenot D., Normand F., 2009)

Le **cycle phénologique** est la succession des stades phénologiques sur un arbre au cours d'une année. Les stades phénologiques de deux cycles de production successifs se superposent dans le temps. Il y a donc, principalement, une différence temporelle entre cycle de production et cycle phénologique.

Nous délimitons le cycle phénologique n entre juin de l'année $n-1$ et juin de l'année n , car juin est un mois de repos végétatif, durant lequel il n'apparaît pas d'activités végétatives ou reproductives en général. Juin peut alors être considéré comme point de départ du cycle de production.

Pour la modélisation, nous nous intéressons à la croissance végétative et à la floraison seulement. La partie fructification sera intégrée au modèle en dehors du cadre de ce stage.

Dans la suite de cette partie, différents concepts biologiques liés à l'architecture du manguiier sont présentées.

1.1.3 Croissance séquentielle et unité de croissance



FIGURE 4 – Une unité de croissance apicale (principale) et deux latérales, (photo Pierre-Eric Lauri)

La croissance du manguiier est rythmique. Un bourgeon débourre (éclo) et donne un axe feuillé appelé **unité de croissance**. Elles sont la base de notre étude et elles sont souvent abrégées en **UC** dans la suite de ce mémoire. Leur durée d'élongation dépend de la température (Dambreville et al., 2013b) et dure en moyenne 10 jours. La longueur de l'unité de croissance ne varie plus ensuite.

Deux mois après, si les conditions sont favorables, les bourgeons de cette nouvelle unité de croissance peuvent débourrer à leur tour et donner de nouvelles unités de croissance. La répétition de ce procédé décrit la **croissance séquentielle**. L'ensemble de l'arbre est ainsi formé par un empilement d'unités de croissance successives.

Une notion de parenté se met en place entre les unités de croissance. Le manguiier peut être vu comme un arbre généalogique. Nous parlons alors d'**UC mère** qui donne des **UC filles**.

1.1.4 Position apicale, latérale et terminale

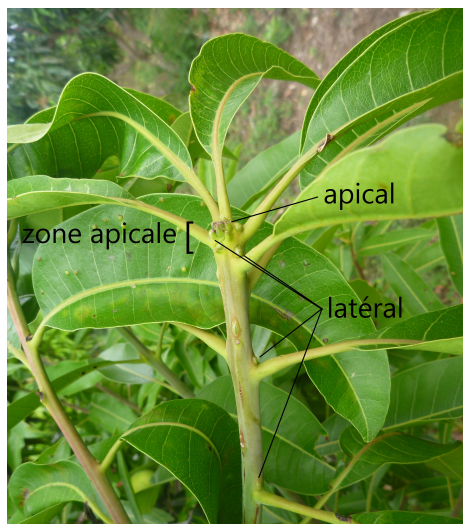


FIGURE 5 – Positionnement des bourgeons le long d'une UC

Une UC, ou un bourgeon, est dit en position **apicale** lorsqu'elle (il) se trouve à l'extrémité de l'unité de croissance.

Une UC, ou un bourgeon, est dit en position **latérale** lorsqu'elle (il) se trouve sur les cotés de l'unité de croissance, au niveau de la base des feuilles (Figure 4 pour les UCs et Figure 5 pour les bourgeons).

Il est important de situer le positionnement d'un bourgeon ou d'une unité de croissance. En effet, la position relative d'une unité de croissance, apicale ou latérale, par rapport à celle qui la précède, a un effet sur la morphologie de cette unité de croissance et sur sa probabilité de croissance végétative, de floraison et de fructification (Normand et al., 2009).

La notion de position terminale s'applique à une autre échelle : l'arbre. Une unité de croissance est dite **terminale** si elle se trouve, à la fin de la saison de croissance végéta-

tive, à l'extrémité de l'arbre. Il peut donc y avoir des unités de croissance terminales en position apicale ou latérale (Figure 4).

Chez le manguier, les bourgeons qui débourrent sont principalement le bourgeon apical et les bourgeons latéraux qui sont au niveau de la zone apicale (extrémité de l'UC). Dans la suite de cette étude, nous supposons que ce sont les seuls bourgeons qui débourrent, car l'étude sur le positionnement des UC latérales, le long de l'axe de l'UC mère, n'a pas été prise en compte.

Notons que ces bourgeons peuvent donner des unités de croissance végétatives ou reproductives (c'est à dire les inflorescences, cf. section suivante). La différenciation des bourgeons est due à des facteurs exogènes (externes à l'arbre), comme la température, et à des facteurs endogènes (internes à l'arbre) comme le diamètre de l'unité de croissance mère, sa position ou sa date d'apparition (Normand et al., 2009 ; Dambreville et al., 2013a). Dans le cadre de ce travail de modélisation, nous considérons uniquement des facteurs endogènes.

1.1.5 Inflorescence



FIGURE 6 – Photo d'une inflorescence

L'**inflorescence** est la partie reproductrice de l'arbre. Les inflorescences sont des structures ramifiées avec une forme le plus souvent pyramidale. Le nombre de fleurs qu'elles contiennent dépend de la variété du manguier. Elles poussent presque exclusivement sur les unités de croissance terminales.

Les inflorescences débourrent au cours de l'hiver austral, c'est-à-dire entre juin et octobre. La floraison est induite par les températures fraîches (Davenport, 2009). De plus, un âge minimal de 7 semaines serait nécessaire pour que l'unité de croissance puisse produire des inflorescences, suggérant la nécessité d'une certaine maturité de l'UC pour fleurir (Nunez-Elisea et Davenport, 1995).

L'inflorescence se dessèche et disparaît à la fin de la floraison, si elle ne donne pas de fruit et après la récolte si elle a donné des fruits.

1.1.6 Nature végétative vs. nature florifère

Une unité de croissance est dite de **nature florifère** si elle a porté au moins une inflorescence. Sinon, elle est dite de **nature végétative**. La nature de l'unité de croissance dépend donc de ce que vont donner ses bourgeons lorsqu'il vont débourrer. Lors de la période de croissance végétative, les UC sont de nature végétative. La distinction entre nature végétative et nature florifère se fait pour les UC terminales, qui sont susceptibles de fleurir, à la fin de la période de floraison.

1.1.7 Croissance séquentielle vs. réitération

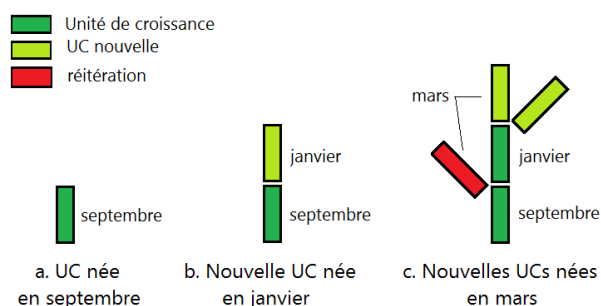


FIGURE 7 – Exemple de croissance séquentielle et d'une réitération

Par exemple, le passage de la figure 7a. à la figure 7b. est une croissance séquentielle. Le passage de la figure 7b. à la figure 7c. présente une croissance séquentielle mais aussi une réitération. Sur la figure 7c., il y a trois nouvelles unités de croissance apparues en mars. Les UC de couleur vert clair sont séquentielles par rapport à leur UC mère apparue en janvier, tandis que l'UC de couleur rouge est une réitération.

Le cas des réitérations n'est pas considéré dans ce premier travail de modélisation car elles ne sont pas fréquentes. Il était néanmoins nécessaire de signaler cette simplification du modèle.

1.1.8 Description schématique de la ramification

Nous avons vu que la croissance séquentielle induit une succession d'unités de croissance. Cette succession se décompose en étages de croissance schématisés sur la figure 8. Chaque segment représente une UC. Nous nommons **UC ancêtre**, les UC terminales à la fin de la croissance végétative d'un cycle de production $n - 1$, et qui sont donc susceptibles de fleurir et fructifier, ou de rester végétatives.

Les UCs ancêtres (étage de croissance 0) sont des UCs mère qui produisent des UCs filles appartenant au 1^{er} étage de croissance du cycle de production n . Ces UCs filles deviennent elles même des UCs mères qui donnent les UCs filles du 2^{ème} étage de croissance, et ainsi de suite jusqu'à la fin de la période de croissance végétative du cycle de production n .

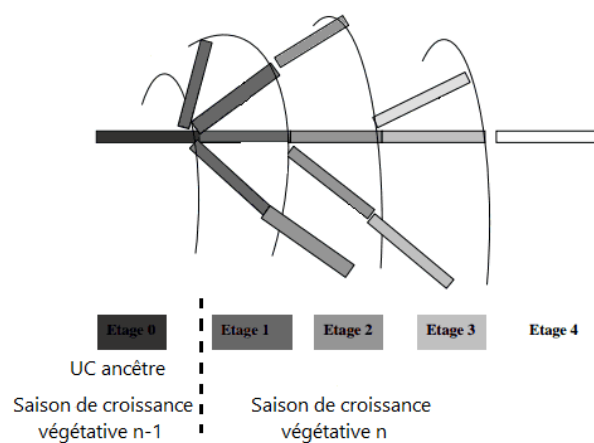


FIGURE 8 – Schéma de ramification à partir d'une UC ancêtre de nature végétative, lors de la saison de croissance végétative (Pambo Bello, 2006)

Un phénomène complémentaire apparaît dans ce schéma de ramification en étages, qui est celui de la ramification d'une UC mère en plusieurs vagues d'apparition de ses UCs

filles. Une UC mère ayant émis des UCs filles, peut à nouveau se ramifier. Cela implique que les UC filles d'une même UC mère peuvent avoir différentes dates d'apparition (d_i sur la Figure 9) lorsqu'il y a eu ramification en plusieurs vagues.

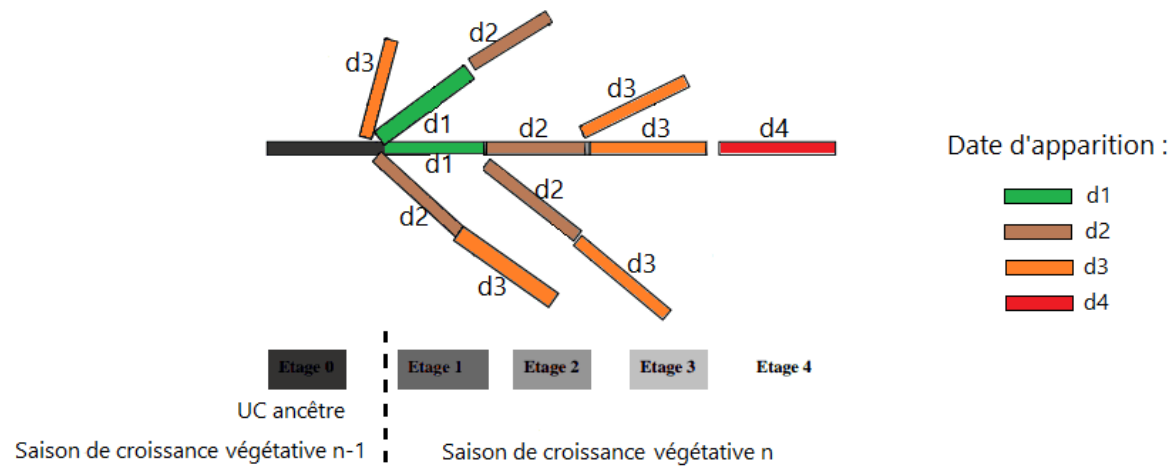


FIGURE 9 – Dates d'apparition (d_i) des UCs issues d'une même UC ancêtre au cours d'un cycle de production, mettant en évidence la ramification des UCs mères en plusieurs vagues d'apparition de leurs filles (Pambo Bello, 2006)

Dans ce stage, la première approche a été de simplifier ce phénomène en supposant qu'il y a eu une seule vague d'apparition des UCs filles pour une UC mère. Nous prenons en compte la première date d'apparition des UC filles. Dans une seconde approche basée sur les modèles graphiques, ce phénomène sera pris en compte.

1.2 Matériels

Afin de modéliser le comportement du manguier, beaucoup de mesures ont été effectuées sur différentes variétés de manguier. Dans cette section, nous présentons les données recueillies et les premières analyses de ces données, effectuées par Abdoul Pambo Bello et Annaëlle Dambreville.

1.2.1 Le verger et les données

Les arbres ont été mis en place en 2001 à St Pierre, Ile de la Réunion. Ce verger est composé de 112 arbres : 8 variétés avec 14 arbres chacune.

L'étude s'est déroulée sur deux cycles de production complets, de juin 2003 à février 2005 pour le cycle 4, et de juin 2004 à février 2006 pour le cycle 5. Sept variétés au port architectural et au comportement agronomique contrastés ont été considérées pour l'acquisition des données. Parmi celles-ci, deux sont cultivées à grande échelle à La Réunion (José et Cogshall) et les autres (Irwin, Kensington Pride, Kent, Nam Doc Mai et Tommy Atkins) font l'objet de cultures à grande échelle dans différents pays producteurs et d'échanges commerciaux internationaux.

Pour chacune des 7 variétés, 5 arbres ont été sélectionnés en mars 2003, avant leur première floraison. Pour évaluer l'influence de la présence de fruits sur la croissance végétative et la floraison au cycle suivant, 2 arbres sur les 5 sélectionnés ont été entièrement éclaircis (retrait total des fruits) après la floraison, dès le début de la nouaison¹ en novembre 2003, puis en octobre 2004, c'est à dire lors de chaque cycle de production étudié. Précisons donc que les arbres éclaircis ne portent pas de fruits, ils sont désignés par le terme "**non chargé** (NC)", les autres arbres ont porté des fruits et sont nommés "**chargé** (C)".

L'étude de ces 35 arbres est faite en décrivant de façon exhaustive la croissance végétative, la floraison et la fructification. Pour chaque arbre, toutes les unités de croissance sont décrites sur la base de caractéristiques quantitatives, qualitatives et temporelles. Les caractéristiques sont les suivantes :

- * Quantitatives :
 - Nombre d'UC filles produites par UC mère
 - Nombre d'inflorescences produites par UC mère
- * Qualitatives :
 - Position relative de l'UC (apicale ou latérale) par rapport à l'UC précédente
 - Nature de l'UC
- * Temporelles :
 - Mois d'apparition de chaque nouvelle UC fille (Pour quatre variétés au cycle 4 et toutes les variétés au cycle 5)
 - Date de pleine floraison de chaque inflorescence (données non complètes pour le cycle 4)

Les données ont été enregistrées dans une base Excel (la "base Doralice"). Une ligne correspond à une unité de croissance, et les colonnes contiennent ses caractéristiques des-

1. transformation de l'ovaire de la fleur fécondée en fruit : début de la croissance des fruits

criptives, phénologiques et de filiation : de quel arbre elle est issue, de quelle charpentière² elle est issue, son UC mère et ses UC filles, sa nature, sa date d'apparition etc.

Lors de ce stage, seule la variété Cogshall a été considérée pour construire les modèles de croissance que nous présentons dans les sections 4 et 5. Ces modèles seront ensuite étendu aux autres variétés en les paramétrant avec leurs données propres.

Chaque unité de croissance possède aussi un code qui permet de la repérer dans le verger, puis dans l'arbre en partant du tronc. Ce code permet de reconstituer l'arbre partiellement car la structure a été décrite à partir de mars 2003. Il manque donc les informations pour "raccorder" ces UCs au tronc. Une autre base de données a complété les informations architecturales afin de reconstituer totalement chaque arbre.

Dans le but de mieux représenter l'architecture des arbres et de pouvoir utiliser des outils appropriés pour analyser des structures arborescentes, les données des deux bases ont été mises sous un format MTG (Multiscale Tree Graph). Nous détaillons ce format dans le chapitre suivant.

Les données recueillies ont déjà fait l'objet de diverses analyses.

1.2.2 Etat de l'art des analyses de la base Doralice

Une première analyse

Une première analyse de la base a été faite en 2006, (Pambo Bello, 2006).

La démarche de modélisation de l'auteur consistait à décomposer la croissance végétative du manguier en processus élémentaires et à les caractériser, puis à rechercher l'effet de différents facteurs sur ces processus.

Pour une UC mère, les processus élémentaires de croissance considérés sont au nombre de 4 :

- démarrage de la croissance ou non (l'UC mère fait-elle au moins une UC fille ?)
- si oui, émet-elle une UC en position apicale ? (Variable importante car affecte l'émission des UC latérales à travers le phénomène de dominance apicale)
- combien d'UC latérales ont été émises ?
- enfin à quelle date apparaissent ces nouvelles UC filles ?

Une schématisation de ces processus élémentaires est présentée sur la figure 10.

2. premier niveau de ramification du tronc

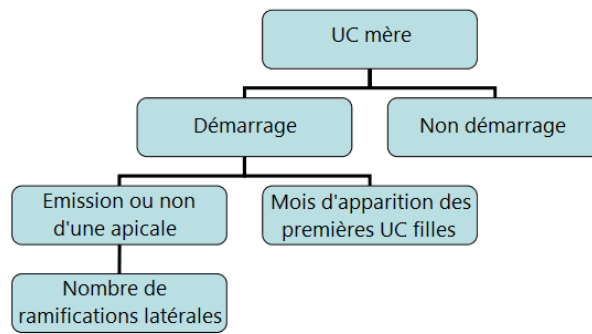


FIGURE 10 – Processus élémentaires de ramification et de croissance à l'échelle d'une UC mère, (Pambo Bello, 2006)

Les processus élémentaires se succèdent et s'enchainent pour former des étages de croissance au cours du cycle de croissance végétative. Son étude a été réalisée séparément pour chaque étage de croissance. Les variables de son modèle décrivent les processus élémentaires de la ramification. Il y a quatre variables :

- la variable aléatoire "démarré", qui représente la probabilité pour l'UC mère, d'émettre au moins une UC fille. C'est-à-dire que c'est une variable binaire valant 1 si l'UC démarre et 0 sinon
- la variable binaire "émission/ non émission d'une UC apicale"
- la variable aléatoire définissant le "nombre d'UC en position latérale émises"
- la variable temporelle "date", qui est le mois d'apparition des UCs filles

La variable du nombre d'UC en position latérale est une variable quantitative dont les réalisations sont le nombre d'UC émises. Un exemple de sa distribution chez la variété Cogshall est :

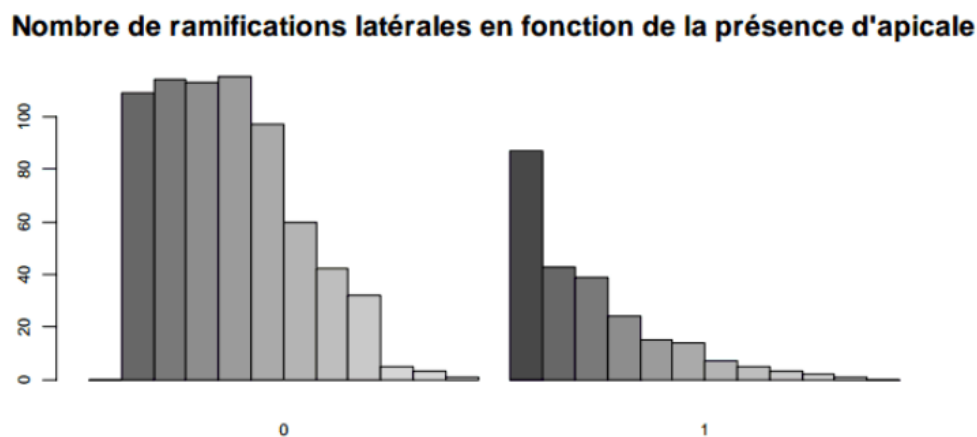


FIGURE 11 – Répartition du nombre de ramifications latérales en fonction de la présence ou non d'une UC apicale, (Pambo Bello, 2006)

Les distribution de la figure 11 peuvent être approchées par une distribution de Poisson.

La variable "date" est une variable aléatoire catégorielle, c'est-à-dire nominale et scindée en K catégories. De plus les K catégories, ici les mois, sont ordonnées et alors la

variable est de distribution multinomiale ordinale.

Son étude s'inscrit dans la recherche des facteurs affectant la croissance végétative du manguiier. Pour cela, il a utilisé des modèles linéaires généralisés (GLM) avec comme variable réponse chacune des quatre variables précédentes. Les facteurs qu'il a retenus sont :

- la nature de l'UC ancêtre (végétative, florifère ou fructifère),
- la position de l'UC ancêtre (apicale ou latérale),
- la charge en fruit de l'arbre (chargé ou non chargé),
- la position de l'UC mère (apicale ou latérale),
- le mois d'apparition de l'UC mère.

Les trois premiers facteurs sont relatifs à la mise en place du premier étage de croissance à partir de l'UC ancêtre. Les deux autres facteurs sont relatifs à la mise en place des étages de croissance 2 et plus.

Les résultats ont permis d'identifier des facteurs affectant la croissance végétative. La nature et la position de l'UC ancêtre et la charge en fruits de l'arbre affectent directement ou indirectement les processus élémentaires de la croissance. Il a aussi observé une régularité et retrouvé une cohérence dans les résultats des analyses pour chaque cycle de production.

Une deuxième analyse

Dans la continuité de ce travail, ces analyses ont été étendues à la reproduction et à d'autres variétés. Elles ont été effectuées par Annëlle Dambreville dans le cadre de sa thèse. Elle a simplifié l'analyse en considérant une seule vague d'UCs filles par UC mère. Son étude a été plus globale dans le sens où elle ne différencie pas les étages de croissance mais les intègre en tant que facteur. Un schéma synthétique de son analyse est présenté sur la figure 12.

Pour un cycle de production, GC2, nous avons un modèle de développement végétatif (A) et un modèle de développement reproductif (B). Les noms encadrés représentent les variables réponses d'un modèle linéaire généralisé (GLM). La forme de l'encadrement des variables indique la distribution de celles-ci. Nous retrouvons les mêmes distributions que pour les analyses précédentes. Les facteurs influençant les variables sont scindés en deux groupes. Il y a des facteurs provenant du cycle phénologique étudié (cadre gris clair : à l'intérieur du cycle) et des facteurs provenant du cycle phénologique précédent (cadre noir : entre les cycles). Pour les GLM, l'interaction d'ordre deux entre les facteurs a été prise en compte.

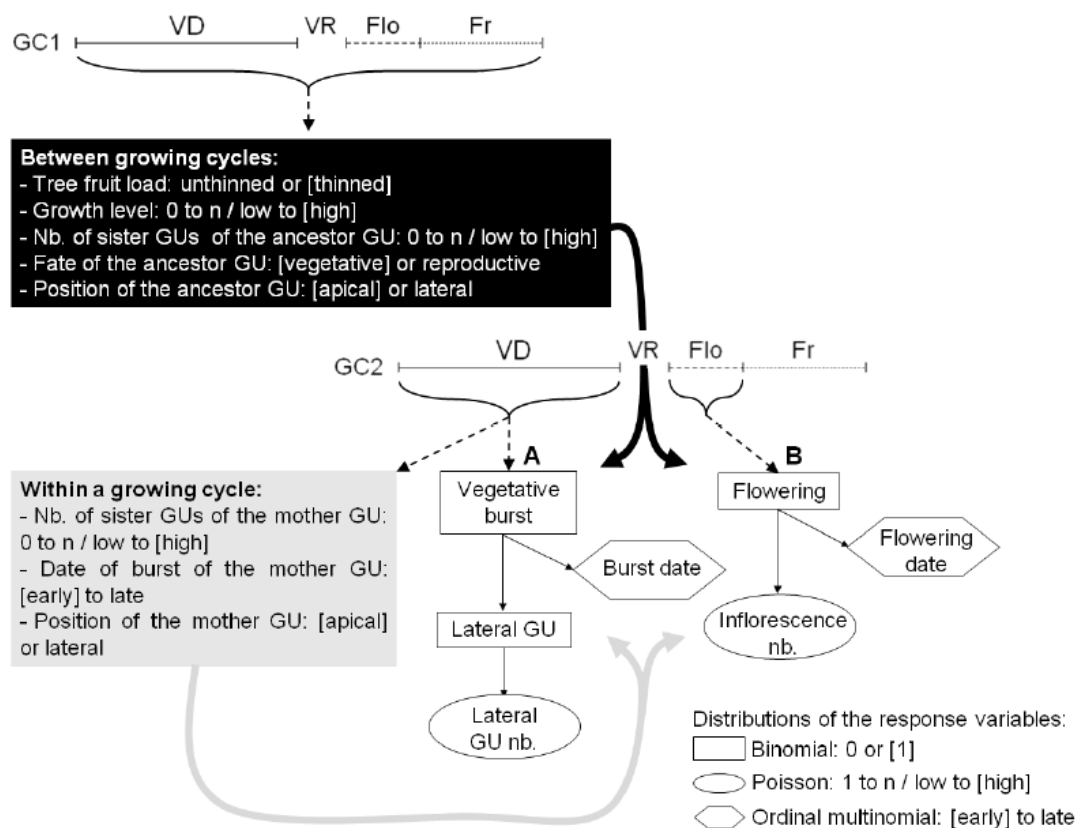


FIGURE 12 – Modèles de développement végétatif (A) et reproductif (B) du manguier, superposé d'une succession temporelle de cycles de croissance (Dambreville et al., 2013a)

Par exemple, les facteurs influençant pour la variété Cogshall sont :

* **A l'intérieur du cycle de production :**

- la date d'apparition de l'UC mère
- la position de l'UC mère

* **Entre les cycles de production :**

- la charge en fruit de l'arbre
- la nature de l'UC ancêtre
- la position de l'ancêtre

Ces modèles de développement végétatif et reproductif ont été appliqués dans le cadre de ce stage. Nous avons également retenu les facteurs influençant des modèles. Cependant, pour la simulation en 3D, nous ne pouvons pas différencier les facteurs propres au cycle de production et les facteurs relatifs au cycle précédent, qui agissent sur la même variable réponse. Nous avons donc été amenés à sélectionner quelques-uns de ces facteurs pour recalculer les GLMs avec ceux-ci (voir section 4) dans l'optique de la modélisation.

2 Le contexte des logiciels

Dans ce rapport, nous cherchons à modéliser la croissance du manguiier en décrivant son architecture. Différents logiciels, pertinents pour notre travail d'analyse et de modélisation, ont été utilisés. Nous les présentons dans cette partie pour donner le cadre informatique de ce travail.

Python



Python est un langage de programmation objet. Il possède un vaste ensemble de bibliothèques scientifiques. Sa syntaxe est caractérisée par la clarté et la concision car l'indentation remplace les accolades.

Ce sera le langage de base pour l'utilisation des autres logiciels.

OpenAlea et VPlants

OpenAlea (Pradal et al., 2008 ; <http://openalea.gforge.inria.fr>) est un projet open source développé principalement pour la communauté de recherche sur les plantes. Il s'agit d'un effort de collaboration pour développer des bibliothèques et des outils de Python, qui répondent aux besoins des travaux actuels et futurs en matière de modélisation de l'architecture des plantes. OpenAlea est une plate-forme contenant des modules qui servent à analyser, visualiser et modéliser le fonctionnement et la croissance de l'architecture des plantes.



VPlants est une suite logicielle rassemblant tous les outils et modules de l'équipe Virtual Plant. Certains sont intégrés à OpenAlea, les autres ont vocation à l'être plus tard lorsqu'ils seront suffisamment finalisés.

Parmi les modules de VPlants que nous utilisons, il y a le module MTG.

Le MTG

Le MTG représente trois notions : un formalisme mathématique (graphe multi-échelles), un format de codage et un module informatique pour manipuler et analyser la structure d'une plante. Suivant le niveau de détail auquel nous considérons une plante, sa géométrie peut être représentée par différents schémas, de complexité variable. La figure 13 montre par exemple différentes échelles de représentation d'un arbre : l'échelle de l'arbre, l'échelle des branches principales (axes), l'échelle des branches secondaires et l'échelle des feuilles.

Afin de prendre en compte ces différentes échelles, les **MTG** (Multiscale Tree Graph) ont été élaborés. Un MTG est constitué d'un ensemble de graphes d'arbre, représentant la topologie de la plante à différentes échelles (tronc, axes, unités de croissance, etc.). A n'importe quelle échelle, les éléments sont reliés par deux types de relations. Ces deux relations correspondent aux deux mécanismes de base de la croissance des plantes, à savoir la croissance apicale et les processus de branchement (croissance latérale). Les relations généalogiques sont maintenues, c'est à dire que les relations parent-enfants sont décrites et nous pouvons retrouver l'ancêtre commun de deux UCs.

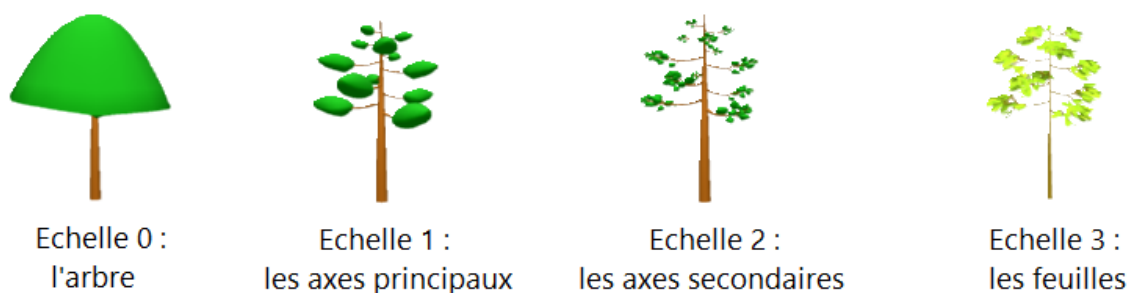


FIGURE 13 – Schéma de représentation d'un arbre à différentes échelles

Différents types d'attributs peuvent être associés aux composants de la plante. Les attributs représentent des propriétés géométriques et morphologiques (exemple : le diamètre d'une tige ou le nombre de fleurs portées).

Le MTG de la base Doralice a été construit à partir des codes de chaque unité de croissance. Ce code indique le positionnement de l'UC dans l'ensemble de la structure décrite pour chaque arbre.

Dans la base de données sous format MTG, les informations ont été récupérées à l'échelle de l'unité de croissance (par exemple, la position de l'UC, la date d'apparition, etc.). Lorsqu'une UC porte une ou plusieurs inflorescences, les inflorescences sont représentées par un unique descendant de l'UC et leurs caractéristiques sont associées à cet unique descendant (par exemple, le nombre d'inflorescences apicale et latérales, la date de pleine floraison).

Le logiciel L-py

L-py (Boudon et al., 2012) est un logiciel qui permet de visualiser en 3D l'architecture et le développement des plantes. Il est basé sur le formalisme des L-systèmes qu'il intègre au langage Python. Ainsi les utilisateurs peuvent donner textuellement des règles de croissances qui seraient exécutées par le logiciel pour générer une forme complexe de la plante.

Ce logiciel est donc central pour la modélisation de la croissance du manguiier.

Le module statistics

Le module statistics complète les outils de statistiques existant dans les bibliothèques de Python. Elle propose un ensemble de procédure statistique : différentes distributions marginales, régressions et estimation de données. C'est à l'aide de ce module que la modélisation, basée sur les modèles graphiques, a pu être effectuée.

Le logiciel R

R est un logiciel libre de traitement des données et d'analyse statistiques. Ce logiciel a été utilisé pour réaliser les analyses avec les modèles linéaires généralisés (GLM).

3 Vers l'architecture du manguiier...

Dans cette partie, nous allons voir comment simuler l'architecture du manguiier. Pour cela nous utilisons le langage des L-systèmes que nous présentons dans une première partie de cette section. Puis nous présentons un premier modèle de croissance du manguiier et enfin, nous expliquons la façon dont nous validons ce modèle par rapport aux données observées.

3.1 Simuler avec le langage des L-systèmes

Le formalisme des L-systèmes a été introduit par Aristid Lindenmayer pour modéliser et simuler le développement d'une plante. Les L-systèmes sont basés sur trois concepts : une chaîne de caractères pour représenter la structure de la plante, une interprétation géométrique de cette chaîne pour visualiser cette structure et des règles de réécriture pour simuler l'évolution de la structure. Pour exprimer tout cela, un langage a été formalisé. Les explications qui vont suivre sont issues de [Boudon, 2004], de [Prusinkiewicz et al., 1997] et de [Boudon, et al., 2012].

3.1.1 La structure

La chaîne de caractères :

Formellement, le langage des L-systèmes est défini à l'aide d'un alphabet :

- Un **alphabet** est un ensemble fini de lettres noté V . Les lettres sont aussi appelées **modules**.

Une branche peut être décrite par des segments de différentes longueurs (Figure 14).

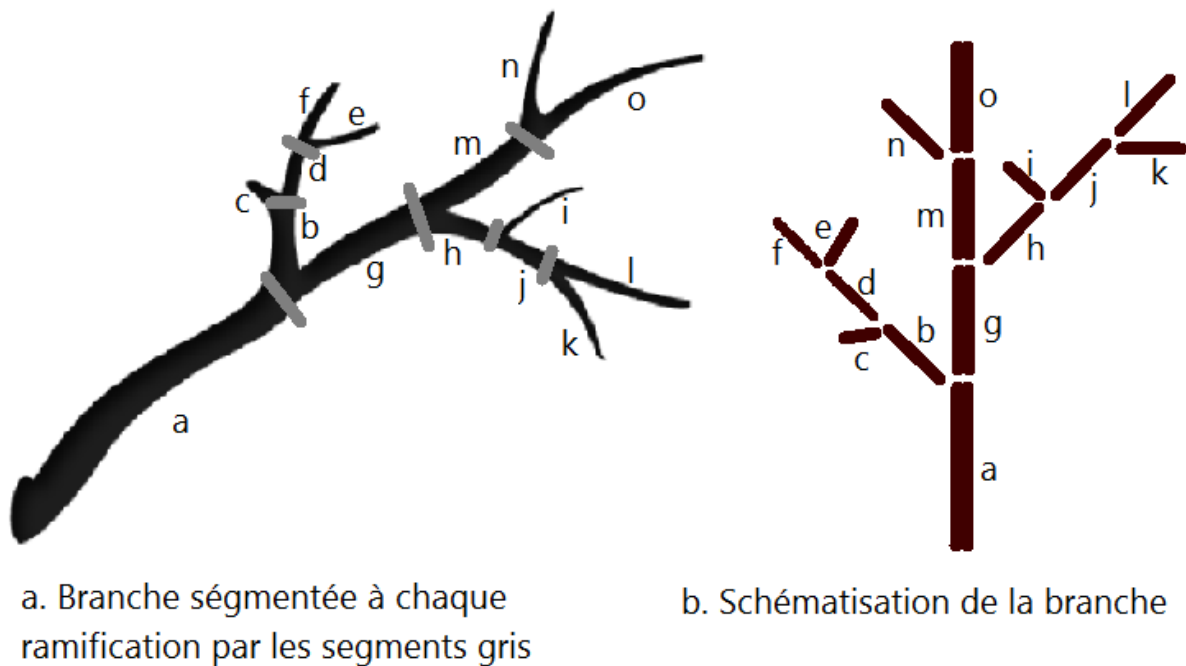


FIGURE 14 – Une branche à une schématisation de sa structure

Ces différents segments peuvent être représentés par des modules (lettres). Une chaîne, composée de plusieurs modules, retranscrit la structure de la branche. Pour cela les modules sont nommés les uns après les autres en partant du bas de la branche et en allant jusqu'aux extrémités. Le début et la fin d'une ramification sont introduits par les symboles "[" et "]". Prenons l'exemple de la figure 14.

La branche est représentée par la chaîne de caractères suivante :

$$branche = a [b [c] d [e] f] g [h [i] j [k] l] m [n] o$$

Les chaînes paramétriques :

Les modules peuvent prendre différents paramètres afin de mieux les caractériser. Par exemple, la longueur du segment est rentrée en paramètre dans la chaîne de caractère :

$$branche = a(2) [b(0.5) [c(0.1)] d(0.5) [e(0.5)] f(0.5)] g(1) \\ [h(0.5) [i(0.2)] j(0.5) [k(0.5)] l(0.5)] m(1) [n(0.5)] o(1)$$

Cette paramétrisation sera pratique pour l'interprétation géométrique.

3.1.2 La géométrie : représentation 3D, la Tortue

Jusqu'ici, nous avons écrit la chaîne de caractères avec des modules et les symboles [et] pour signifier une ramification. Mais cela n'est pas suffisant pour décrire fidèlement la géométrie d'une plante (par exemple, l'angle entre un segment et une ramification n'est pas donné). C'est pourquoi certains symboles, avec ou sans paramètres, ont été introduits pour représenter des propriétés géométriques de la structure décrite. Pour calculer une géométrie, la chaîne à interpréter est scannée séquentiellement de gauche à droite, et ces symboles consécutifs sont interprétés comme des commandes qui manœuvrent une **tortue**.

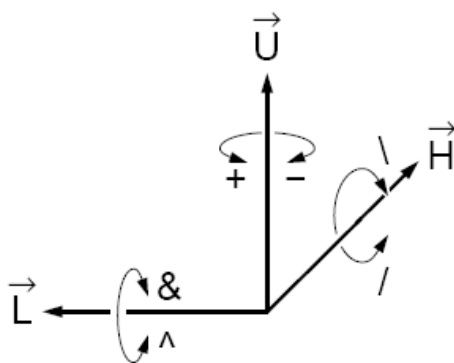


FIGURE 15 – Le repère de la tortue, (Heading, Left, Up) avec les symboles de rotation

La Tortue est un concept représenté par un repère mobile (Figure 15) et possédant un certain nombre de propriétés permettant de tracer les figures géométriques.

La tortue est représentée par son état, qui consiste en une position et une orientation dans le système de coordonnées cartésien. Elle possède des attributs supplémentaires comme la couleur courante et l'épaisseur de trait. La position est définie par un vecteur P , et l'orientation par trois vecteurs H , L et U , indiquant la direction (Heading), la gauche (Left) et le haut (Up).

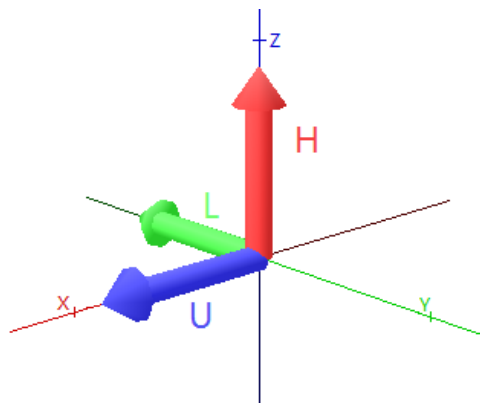


FIGURE 16 – Le repère de la tortue dans L-py

Dans L-py, la tortue est initialement située à l'origine du système de coordonnées cartésien, avec le vecteur de direction H pointant dans la direction positive de l'axe des z , et le vecteur gauche L pointant dans la direction négative de l'axe y (Figure 16).

Les actions de la tortue et ses changements d'états sont causés par l'interprétation de symboles spécifiques (par exemple $+$, $-$, $\&$, \wedge), chacun pouvant prendre des paramètres (par exemple l'angle de rotation, la longueur du segment à tracer). Si le symbole n'est suivi par aucun paramètre, des valeurs par défaut sont utilisées.

Un exemple d'interprétation par la tortue d'une chaîne de caractères est donné sur la figure 17. Les symboles F sont interprétés par le tracé d'un cylindre, la longueur et la largeur de ce cylindre étant par défaut égal à 1 et 0.1 respectivement. La couleur par défaut est le marron. Les symboles $-$ et \backslash correspondent respectivement à des rotations de la tortue autour des axes U et H . Lorsqu'elle lit le symbole $[$, la tortue enregistre son état en mémoire et le restaure avec le symbole $]$.

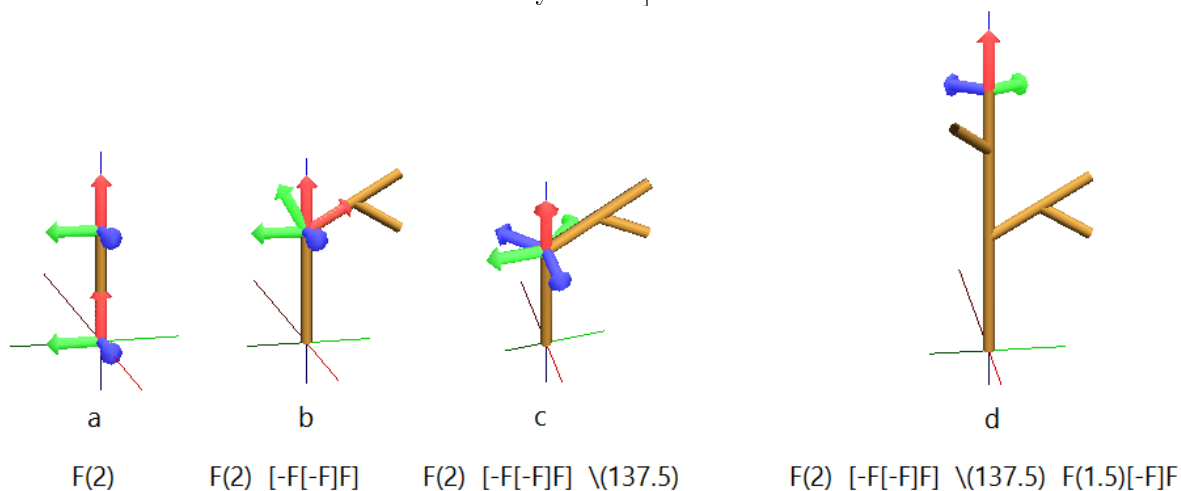


FIGURE 17 – Étapes d'interprétation géométrique de la chaîne.

L'étape a) revient au tracé d'un cylindre de longueur 2. L'étape b) montre la rotation de l'axe U pour tracer la ramification avec un angle par défaut de 60° . L'étape c) indique le retour à la position apicale puis une rotation de l'axe H de 137.5° . Enfin, l'étape d) montre la position de la tortue à la fin du tracé. Une grande partie des symboles interprétés par la tortue se trouvent en annexe A page ?? . Ils s'obtiennent facilement dans l'aide du logiciel L-py.

3.1.3 La dynamique : évolution de la chaîne de caractères

Maintenant qu'une plante peut être représentée par une chaîne de caractères, il est intéressant de faire évoluer la chaîne de caractères pour simuler l'évolution de la plante (ajout de nouvelles pousses, épaississement des axes, ...).

Les L-systèmes :

Différentes définitions sont nécessaires à la compréhension du système de réécriture de la chaîne de caractères.

- Un **alphabet** est un ensemble fini de lettres noté V . Les lettres sont aussi appelées **modules**
- Un **mot** est une séquence de lettres sur un alphabet. L'ensemble de tous les mots sur l'alphabet V est noté V^* .
- Une **production** est une paire (a, u) notée $a \rightarrow u$, où a est une lettre et u est un mot ; a est appelé **prédécesseur** et u est appelé **successeur**.
- Un L-système est un triplet $G = \langle V, \omega, P \rangle$, où V est un alphabet, $\omega \in V^*$ est un mot appelé **l'axiome**, et P est l'ensemble des productions tel que $\forall a \in V : \exists ! p_a \in P$, où p_a représente une production qui a un module a comme prédécesseur.

Prenons un exemple simple pour illustrer les définitions. Soit un axiome ω et deux productions p_1 et p_2 . A gauche nous avons les règles de productions et à droite les étapes du développement :

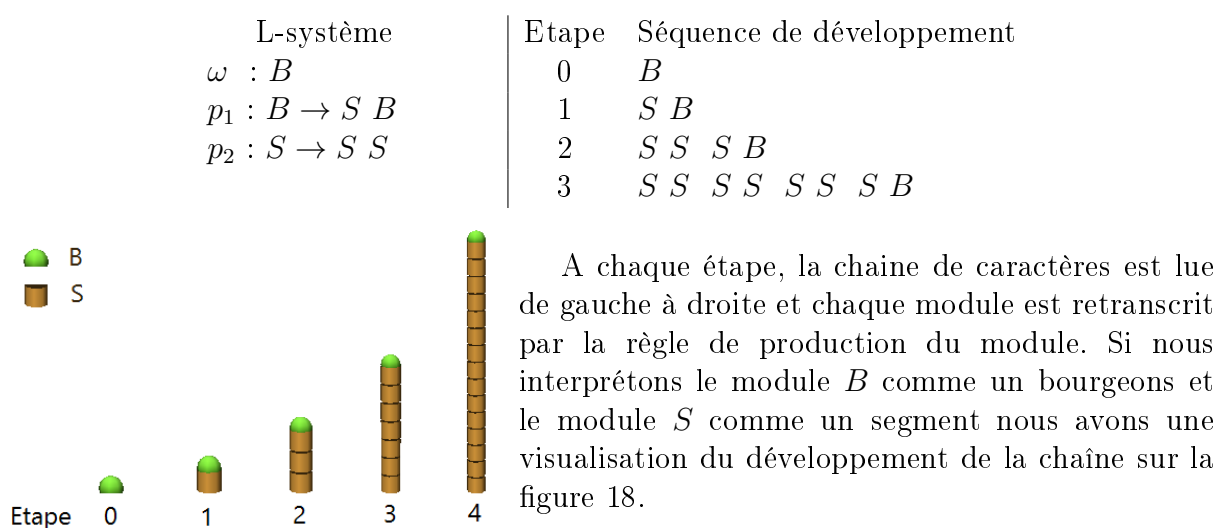


FIGURE 18 – Exemple visuel du développement de la chaîne, à l'aide de L-py

Dans L-py, l'écriture de la règle s'effectue de la manière suivante : nous commençons par déclarer les modules, puis l'axiome et ensuite la règle de production. L'écriture sous L-py est de la forme :

```

1 module S, B
2 axiom : B
3
4 production :
5
6 B:
7   produce SB
8
9 S :
10  produce SS

```

Les lignes 6 et 7 correspondent à la règle $B \rightarrow S B$. Pour produire le successeur, nous utilisons la commande **produce**. La commande **nproduce** permet aussi de produire le successeur en plusieurs étapes.

Les L-systèmes conditionnés :

Il peut être intéressant qu'une production s'applique en fonction d'une condition sur la valeur des paramètres du prédécesseur. La règle s'écrit de la manière suivante :

$$\text{prédécesseur} : \text{condition} \rightarrow \text{successeur}$$

Illustrons la notion en reprenant l'exemple précédent et en ajoutant un paramètre d'âge t . Supposons qu'au bout de deux mois, le bourgeon apical meurt et qu'un bourgeons latéral reprend la suite.

L-système		Etape	Séquence de développement
$\omega :$	B	0	$B(0)$
$p_3 :$	$B(t) : t \geq 2 \rightarrow [B(0)]$	1	$S(1) B(1)$
$p_1 :$	$B(t) \rightarrow S(1) B(t+1)$	2	$S(2) S(1) B(2)$
$p_2 :$	$S(t) \rightarrow S(t+1)$	3	$S(3) S(2) [B(0)]$
		4	$S(4) S(3) [S(1) B(1)]$

Lorsque aucune règle n'est précisé pour un module, la règle Identité s'applique par défaut ($A \rightarrow A$). De même, il y a des règles implicites telles que $[\rightarrow [$ et $] \rightarrow]$. Pour supprimer un module, le symbole $*$ doit être utilisé ($A \rightarrow *$).

La visualisation du développement de la chaîne de caractères est sur la figure 19 :

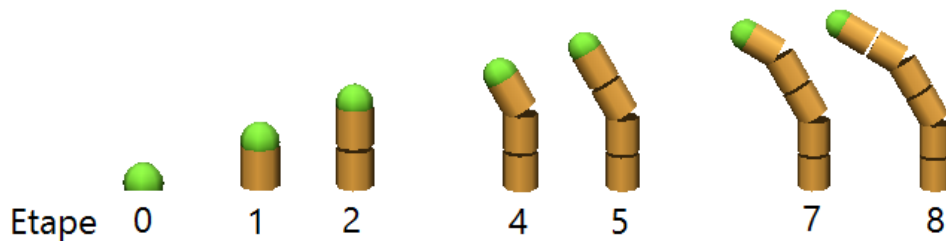


FIGURE 19 – Exemple visuel du développement d'une chaîne dépendant du temps, à l'aide de L-py

Dans L-py, il est possible d'utiliser le langage Python pour créer des conditions. Ainsi la production p_3 s'écrit :

```
B(t) :
  if t >= 2 :
    produce [B(0)]
```

Les L-systèmes sensibles au contexte :

Dans les L-systèmes, l'application des productions peut dépendre aussi du contexte du prédécesseur. Cette extension est nécessaire pour modéliser l'échange d'informations entre modules voisins, pour être capable, par exemple, de simuler le flux d'eau, d'hormones ou de nutriments à travers la plante. Dans notre cas, pour simuler le débourrement d'un bourgeons en fonction des unités mères ou ancêtre et les bourgeons filles. La règle s'écrit de la manière suivante :

contexte gauche \langle *prédécesseur* \rangle *contexte droit* : *condition* \rightarrow *successeur*

Un exemple de règle serait $A \langle B \rangle C \rightarrow D$. Ainsi dans la chaîne de caractères $FABCBC$, le changement s'effectue sur le premier B . Le deuxième B n'est pas concerné par cette règle. Le résultat est $FABCBC \rightarrow FADCBC$.

Les règles de d'interprétation

Nous avons vu qu'il est possible de simuler l'évolution de la structure d'une plante en utilisant des règles de production. L'interprétation géométrique de cette chaîne peut se faire à l'aide des règles spécifiques, dites d'interprétation. Ces règles vont transformer les modules de la chaîne représentant la plante en modules interprétables par la tortue. Ainsi la tortue sera capable de représenter n'importe quel modules. Ces règles sont uniquement appelées pour afficher la structure

Ainsi, les règles d'interprétation pour l'exemple de la figure 19 sont :

interpretation:

$B(t) \rightarrow , (green) @0$

$S \rightarrow , (brown) F(0.5)$

Modélisation géométrique des organes d'une plante

Il est connu que les plantes sont attirées par le soleil, c'est pourquoi elles peuvent adopter une forme particulière pour capter les rayons lumineux. Dans le but de modéliser ce phénomène d'attraction, le **tropisme** a été mis en place. Il faut pour cela indiquer une direction et l'**élasticité** (flexibilité) de la branche. La figure 20 ci-dessous permet de visualiser une branche tracée sans et avec tropisme.

L'élasticité est un nombre compris entre 0 et 1. Lorsque la tortue lit le symbole du tropisme ($@Tp$), elle trace un segment puis croit d'un certain angle α vers la direction

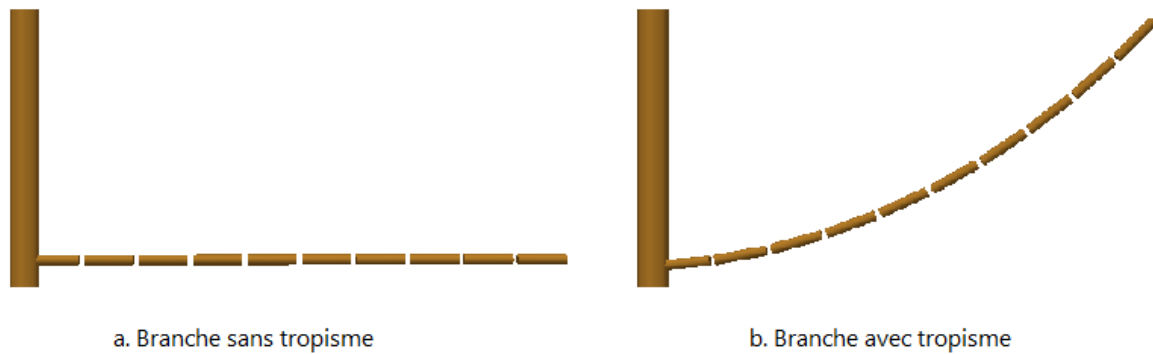


FIGURE 20 – Visualisation de la branche sans tropisme et avec tropisme (vecteur directeur : $(0,0,1)$, élasticité : 0.02)

de tropisme. Le calcul de cet angle se fait à l'aide de l'élasticité *elasticity* : soit θ l'angle entre le vecteur de départ (vecteur H) et le vecteur directeur ; $\alpha = \theta \cdot elasticity$. Cet angle est recalculé à chaque déplacement.

Remarquons que le tropisme peut aussi servir à modéliser l'effet du poids des fruits sur une branche (gravité).

Si la forme de la branche est plus complexe, il est possible de la représenter par une courbe et de faire tracer la tortue en suivant le chemin de la courbe. Sur la figure 21 ci-dessous, voici un résultat de la méthode.

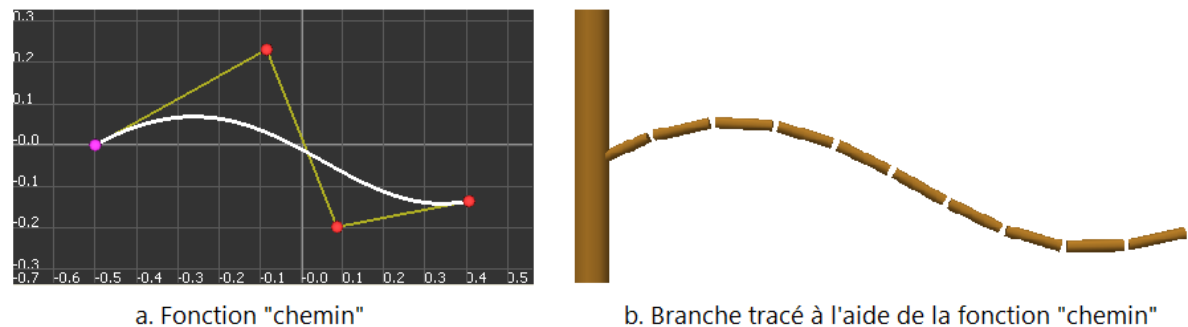


FIGURE 21 – Une branche prenant la même forme que la courbe

Il est aussi possible de dessiner des surfaces.

La fonction *Sweep()* de L-py permet de définir la surface voulue. Elle prend en paramètre 3 courbes et un paramètre de taille :

- une courbe *chemin*, pour tracer la surface
- une courbe *section* qui donne l'allure de la surface
- une fonction *largeur*, la largeur de la section
- une longueur, la longueur du chemin

Par exemple, pour avoir une feuille d'un manguier, nous définissons les courbes *leafpath*, *leafsection* et *leafdiam*. Sur la figure 22, nous avons leurs allures respectives. Pour chaque courbe, la feuille est mise sous un certain angle afin de pouvoir comparer la courbe et le tracé.

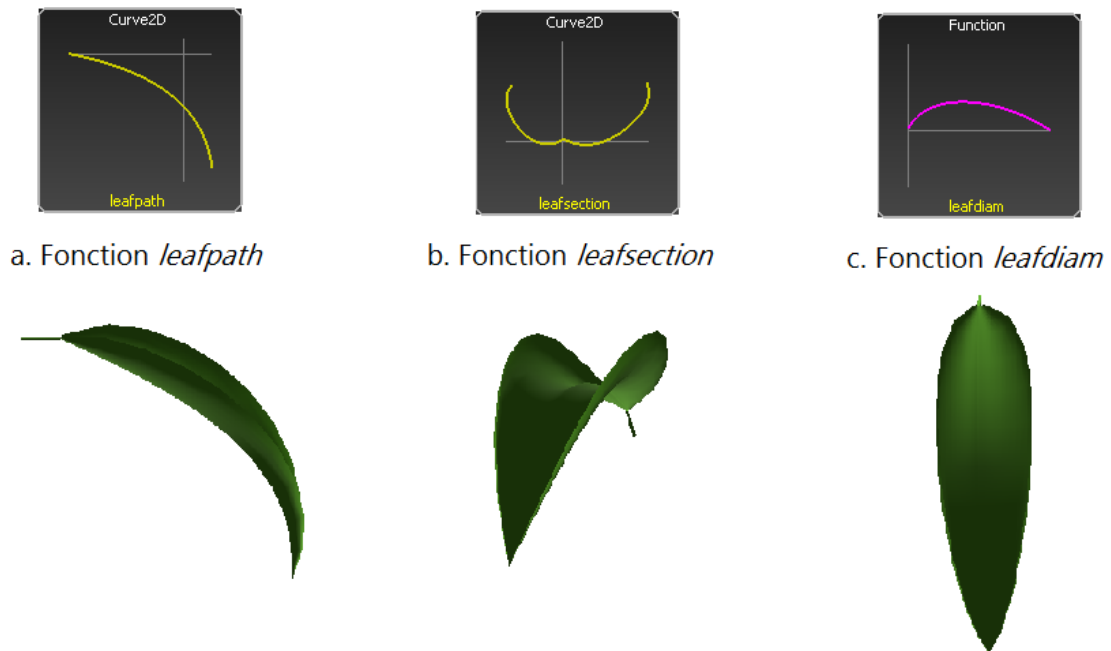


FIGURE 22 – Les différentes fonctions utiles pour tracer une feuille

Le code sous L-py est le suivant :

```
Leaf(length) :
  produce [ &(90) F(0.5,0.01) Sweep(leafpath, leafsection, length,
    leafdiam) ]
```

3.2 Un modèle simple de croissance du manguier

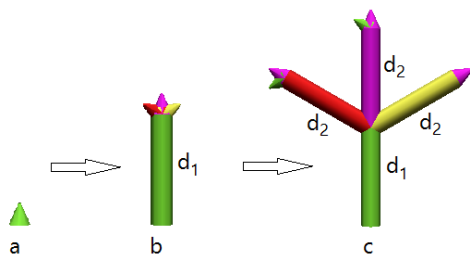


FIGURE 23 – Placement des bourgeons à chaque étape, une étape est représentée par la flèche

La croissance végétative s'effectue par la règle de production suivante : un bourgeon mère³ grandit à une date d_1 pour donner une UC mère. Puis elle place ses bourgeons filles qui vont se développer à une date d_2 et sont positionnées au bout de l'UC. La figure 23 présente deux pas de croissance : le bourgeon vert grandit et place ses enfants, puis les bourgeons filles grandissent et placent leurs enfants respectifs.

3. nous gardons le terme mère, utilisé pour les UCs, au lieu d'écrire père pour ne pas perdre le lecteur

Dans le déroulement de la règle, plusieurs questions se posent :

- combien de bourgeons filles vont apparaître ?
- à quelle date elles grandissent ?
- où les placer ?

Un premier modèle simple de croissance végétative a été fait à partir des connaissances biologiques de base sur le développement du manguiier :

- une UC mère, lorsqu'elle débouffe, peut produire entre 1 et 6 UCs filles (mais 5 ou 6 UCs filles par UC mère sont des cas rares)
- s'il y a au moins une UC fille, la position apicale est obligatoirement prise
- l'écart entre la date de sortie d'une UC mère et la date de sortie des premières UCs filles est d'environ deux mois (± 1 mois)
- pour simplifier le modèle, les UCs filles débouffent au même moment.

La simulation se déroule dans le temps, l'échelle choisie est la semaine. Dans ce premier modèle, nous déterminons le nombre de bourgeons filles et leur date d'apparition de manière arbitraire. La règle est la suivante :

```
1 A(burst_date) :
2     if burst_date <= current_date :
3         burst_date = current_date
4         nproduce UC(burst_date)
5         nb_children = randint(1,4)
6         delta_date_children = randint(4,12)
7         if nb_children > 1:
8             angle_between_children = 360./(nb_children-1)
9             for i in xrange(nb_children-1):
10                 nproduce /( angle_between_children ) [ &(60) A( burst_date +
11                     timedelta(weeks = delta_date_children) ) ]
12             produce A( burst_date + timedelta(weeks = delta_date_children) )
```

Le module A représente un bourgeon, son paramètre (ligne 1) est sa date d'apparition pour la simulation. Lorsqu'il est temps pour le bourgeon de grandir (ligne 2-3), nous remplaçons le bourgeon par une UC possédant les mêmes paramètres que le bourgeon (ligne 4). Ensuite, le nombre de bourgeons filles est déterminé en tirant de façon uniforme un nombre entre 1 et 4 (enfants) (ligne 5). A la ligne 6, nous calculons la différence de temps entre l'apparition de la mère et de ses filles. Cette différence de temps *delta_date_children* est tiré de façon uniforme entre 4 et 12 (semaines). Les lignes de 7 à 13 place les bourgeons filles. Il faut tout d'abord placer les bourgeons latéraux s'il y en a (ligne 7). Les bourgeons latéraux sont placés uniformément autour de l'UC mère, pour cela un angle entre les bourgeons latéraux est calculé (ligne 8). Puis pour chaque bourgeon latéral (ligne 9), la tortue fait une rotation autour de l'UC mère d'un angle *angle_between_children* puis place le bourgeon latéral. La date de débouffement du bourgeons est calculée aux lignes 10-11. Pour finir, nous plaçons le bourgeon apical à la ligne 12.

L'étape suivante est l'interprétation géométrique du module UC. Géométriquement, l'UC a une longueur prédéfinie, une diamètre variable en fonction de son âge. Elle possède des feuilles le long de son axe et plusieurs feuilles au niveau de la zone apicale. Le place-

ment des feuilles le long de l'UC se fait en suivant une spirale. La longueur des feuilles est variable le long de l'UC.

Pour la mise en place du tropisme, l'UC est segmentée en 10 petits segments. L'ajout d'une feuille se fait entre ces petits segments. Voici la règle d'interprétation géométrique :

```

1 total_weeks = 60
2 UC(burst_date) :
3     week_age = (current_date - burst_date).days/7
4     l_radius = radius(1.-(week_age/float(total_weeks)))
5     dl = 1/10.
6     rotation_angle = 137.5
7     for i in xrange(0,10):
8         nproduce F(dl,l_radius) Leaf(rotation_angle*i,l_radius,
9             leafsize(0.1+0.1*i))
10
11 Leaf(angle,radius_uc,length, dl) :
12     produce [ /(angle) &(90) f(radius_uc) Sweep(leafpath,
13         leafsection,length,length*dl,leafdiam) ]

```

Les lignes 2 et 3 calculent l'âge de l'UC en semaine et fournit la longueur du rayon du cylindre avec la fonction *radius* (Figure 24). La ligne 5 indique la longueur des petits segments. *rotation_angle* concerne l'angle de rotation autour de l'UC pour placer une feuille. La valeur de l'angle est prise arbitrairement. Pour chaque petit segment (ligne 7), un cylindre de longueur *dl* et de rayon *l_radius* est tracé, puis une feuille, *Leaf*, est placée.

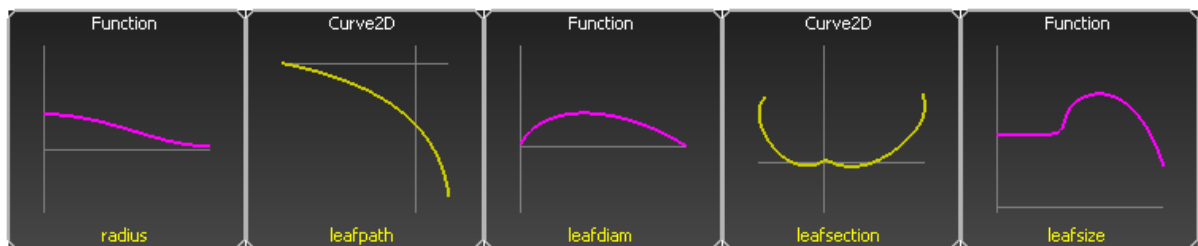


FIGURE 24 – Allure des fonctions utilisées pour la mise en place de la géométrie

La feuille, *Leaf*, est placée en faisant une rotation autour de l'UC. Un décalage horizontal du pétiole (la tige) de la feuille, pour le placer sur le bord du cylindre, est obtenu par une inclinaison de 90° et un déplacement sur le bord de l'UC sans tracer de cylindre (symbole f). La taille des feuilles de longueur de l'axe varie et est modélisée par la fonction *leafsize*. L'allure des fonctions utilisées est présentée sur la Figure 24. La représentation géométrique d'une UC est sur la Figure 25.



FIGURE 25 – Visualisation d'une UC sous L-py

Nous avons ensuite intégré les inflorescences à la règle de production et trois questions sont survenues :

- combien d'inflorescences vont apparaître par UC ?
- à quelle date elles apparaissent ?
- où les placer ?

Pour une première simulation, le principe de la règle est qu'un bourgeon donne une UC ou une inflorescence. L'apparition de l'inflorescence est conditionnée par un facteur exogène, la température moyenne dans le mois, et par un facteur endogène, une probabilité de floraison. Nous savons que les inflorescences apparaissent au cours de l'hiver austral, il faut donc une température moyenne qui soit en-dessous d'un seuil. La règle de production est la suivante :

```

1 P_FLOWERING = 0.35
2 A(burst_date) :
3     if burst_date <= current_date :
4         burst_date = current_date
5         if current_temperature < 20. and uniform(0,1) < P_FLOWERING:
6             nproduce Inflorescence(burst_date)
7         else :
8             nproduce UC(burst_date)
9             nb_children = randint(1,4)
10            delta_date_children = randint(4,12)
11            if nb_children > 1:
12                angle_between_children = 360./(nb_children-1)
13                for i in xrange(nb_children-1):
14                    nproduce /( angle_between_children ) [ &(amp;60) A( burst_date
15                        + timedelta(weeks = delta_date_children) ) ]
16                produce A( burst_date + timedelta(weeks = delta_date_children) )

```

La règle est la même que précédemment pour la croissance végétative. Les changements sont aux lignes 5-6, une inflorescence est produite en position apicale si la température est en dessous de 20°C avec une probabilité de fleurir de 0.35.

Au niveau représentation géométrique, l'inflorescence est un segment fin possédant de nombreuses ramifications et dont la forme est pyramidale. Elle porte de petites fleurs

symbolisées par des sphères. La modélisation de la croissance de l'inflorescence et de la croissance de ses sphères se fait en fonction de son âge. Pour modéliser le poids des futurs fruits, nous avons placé un tropisme vers le sol en fonction de l'âge. C'est en adéquation avec la croissance des sphères. La règle d'interprétation géométrique de l'inflorescence est la suivante :

```

1 Inflorescence(burst_date) :
2     day_age = (current_date - burst_date).days
3     growth_duration = 21.
4     flower_growth_duration = 21.
5     flower_radius = 0 if day_age < growth_duration else
6         0.04*min(1, (day_age-flower_growth_duration)/flower_growth_duration)
7     nproduce [
8         if day_age > growth_duration : nproduce @Tp(0,0,-1)
9             @Ts(inflo_elasticity(min(1, (day_age-growth_duration)/40.)))
10    nproduce , (rouge) _(0.01) F(0.1)
11    nbflower = 4
12    for i in xrange(0,10):
13        nproduce F(0.1) [ /(rd.randint(0,180))
14        for j in xrange(0,nbflower):
15            nproduce [ /(j*90) &(rd.randint(70,110)) F(inflo_radius(i*0.1))
16            if flower_radius > 0: nproduce , (jaune) @O(flower_radius)
17            nproduce ]
18        nproduce ]
19    nproduce , (rouge) F(0.1)
20    if flower_radius > 0: nproduce , (jaune) @O(0.04)
21    nproduce ]

```

Le principe est similaire à la géométrie de l'UC. Le long de l'inflorescence, il y a une segmentation de 10 petits cylindres. Entre chaque cylindre, 4 ramifications de longueur variable sont ajoutées (ligne 14-15). La variabilité de la longueur est due à la fonction *inflo_radius()* (Figure 26). Une sphère est placée au bout de chaque ramification, son rayon dépend de l'âge de l'inflorescence. Le tropisme est placé au début de l'interprétation à la ligne 9. Son élasticité varie en fonction de l'âge. Plus l'inflorescence est mature et plus elle est orientée vers le sol (la gravité).

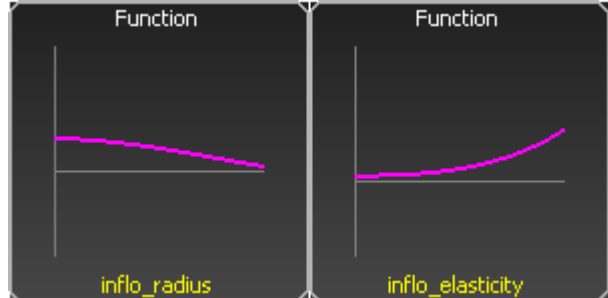


FIGURE 26 – Allure des fonctions utilisées pour la représentation géométrique d'une inflorescence

La croissance d'une inflorescence est présentée sur la Figure 27.

Nous avons ainsi un premier modèle de croissance du manguier. Un exemple de simulation est sur la figure 28.

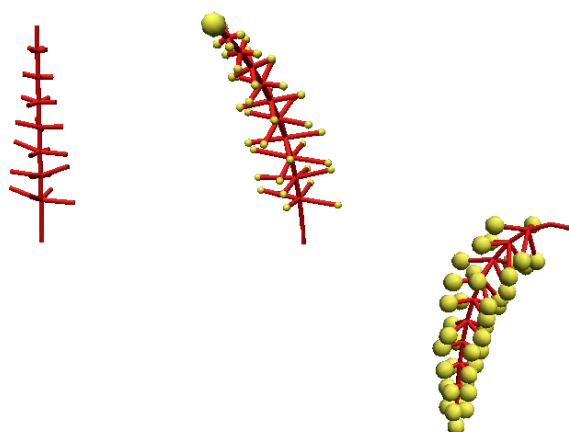


FIGURE 27 – Séquence de croissance d’une inflorescence, le poids des fleurs incline l’inflorescence vers le sol

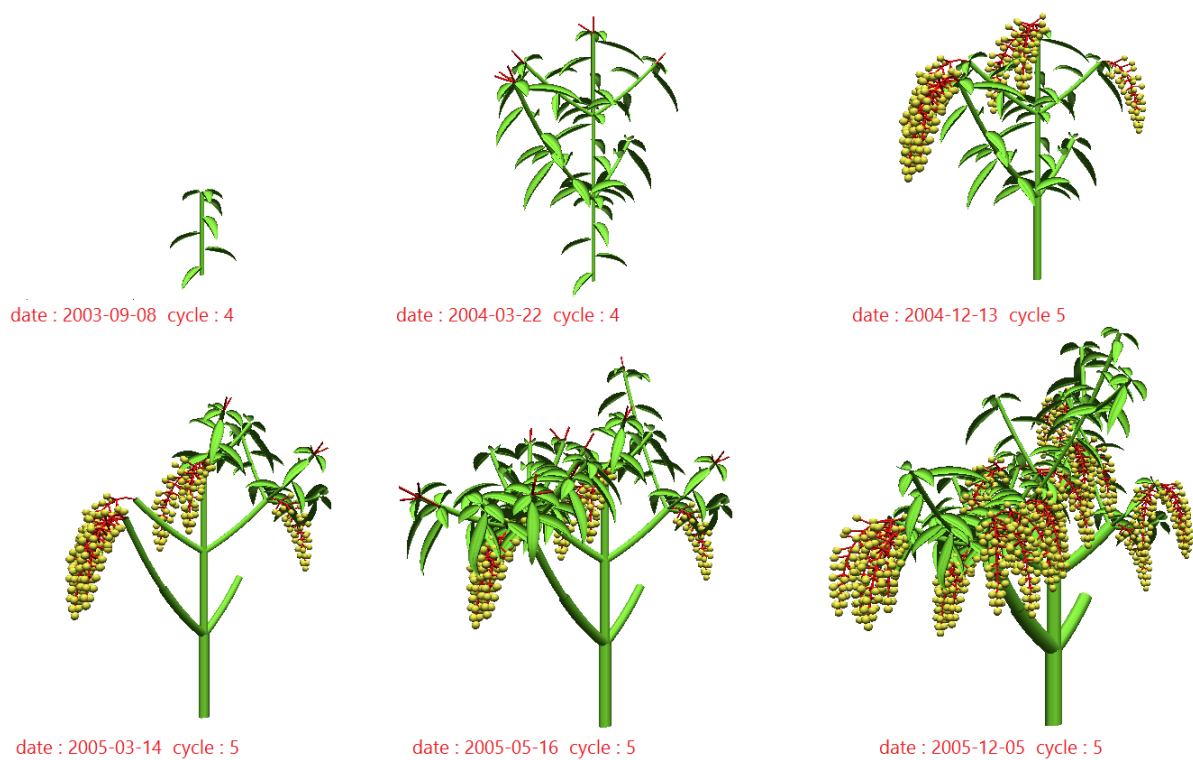


FIGURE 28 – Modélisation de la croissance du manguier

4 Modélisation basée sur les modèles linéaires généralisés (GLM)

Dans la modélisation précédente de la croissance du manguier, l'apparition de nouveaux organes (morphogénèse) est simulée par des règles très simples. Or ce processus est fondamental pour la construction de l'architecture du manguier. Les analyses précédentes nous montrent le lien entre ce processus et différents facteurs structurels et environnementaux. Dans cette section, nous utilisons les modèles linéaires généralisés pour caractériser statistiquement la mise en place des organes en fonction d'une liste de facteurs sélectionnés dans les analyses précédentes. Ces GLM nous donnent la probabilité d'apparition des organes qui ont été intégrés à la simulation. Dans ce chapitre, nous décrivons les GLM, puis nous présentons le processus de croissance qui utilise ces modèles. Enfin nous abordons les procédés de simulation.

Pour la simulation, nous avons utilisé des fonctions de R faisant références aux modèles linéaires généralisés. Afin de mieux comprendre leur utilisation, il est nécessaire de connaître la théorie qui est utilisée. La section qui suit présente les modèles linéaires généralisés.

4.1 Le modèle statistique

La section qui suit est inspirée d'une thèse en cours de rédaction dont l'auteur est Jean Peyhardi.

Nous avons des variables explicatives $X = (X_1, \dots, X_K)$ qui influencent une variable réponse Y , le but du modèle mathématique est de déterminer une fonction f telle que $Y = f(X)$.

Le modèle linéaire

Le modèle le plus simple est le modèle linéaire (f est une fonction linéaire) et s'écrit sous la forme :

$$Y = Z\beta + E$$

où Y est le vecteur réponse (supposé suivre une loi normale $\mathcal{N}(\mu, \sigma^2)$),

B est le vecteur des paramètres que nous cherchons à estimer (l'intercepte et le poids de chaque variable explicative sur Y),

Z est la **matrice de design**,

et E est le vecteur des erreurs (supposé suivre une loi normale $\mathcal{N}(0, \sigma^2)$).

Attention, la notation du modèle efface une information importante : nous désignons par Y la variable réponse mais il s'agit de la variable $Y|X$ qui nous intéresse et qui suit une loi normale ($\forall i \in \{1, \dots, n\}$, $Y_i|X_i = x_i \stackrel{\text{i.i.d}}{\sim} \mathcal{N}(\mu_i, \sigma^2)$, où μ_i est défini à l'aide de x_i et σ^2 est inconnu).

La formulation $Y|X = Z\beta + E$ peut se réécrire : pour tout $i \in \{1, \dots, n\}$, $Y_i|X_i = z_i\beta + \epsilon_i$. Ce qui revient à $\mathbb{E}[Y_i|X_i = x_i] = z_i\beta$ (puisque $\mathbb{E}[\epsilon_i] = 0$).

La matrice de design

Dans le cas où toutes les variables explicatives sont continues, une ligne de la matrice de design s'écrit $z_i = (1, x_{i1}, \dots, x_{iK})$. Sinon, soit k tel que X_k est une variable qualitative et il faut transformer les valeurs x_{ik} par un vecteur ligne suivant :

Si x_k possède M modalités, $x_{ik} = m$ se transforme en

$$x_{ik} = (0, \dots, 0, \underbrace{1}_m, 0, \dots, 0), \text{ pour } m \in \{1, \dots, M-1\}.$$

Si la modalité M est prise, x_{ik} est le vecteur nul de taille $M-1$.

Notons que le produit $\eta_i = z_i \beta$ est le **prédicteur linéaire** de $Y_i|X_i$. Lorsque le vecteur β a été estimé nous obtenons la valeur prédite de $Y_i|X_i$ à une erreur près ϵ_i . L'estimation du vecteur des coefficients β est obtenue en maximisant la vraisemblance.

Les modèles linéaires généralisés (GLM)

Le cas du modèle linéaire permet de prédire les valeurs de Y facilement mais demande de grandes contraintes telles que l'erreur et la variable réponse Y suivent une loi normale. C'est pourquoi les **modèles linéaires généralisés** (GLM pour Generalized Linear Model) ont été introduit. L'écriture du modèle est la suivante, $\forall i \in \{1, \dots, n\}$,

$$\mathbb{E}[Y_i|X_i] = g^{-1}(z_i \beta)$$

où les $Y_i|X_i$ suivent une loi de famille exponentielle,

g est la **fonction de lien**,

z_i est la ligne i de la matrice de design

et β est le vecteur des coefficients du modèle.

Dans ce modèle, la contrainte s'allège dans le sens où il faut que $Y|X$ suive une loi de famille exponentielle. Le choix de cette famille est dû au fait qu'elle possède de bonnes propriétés pour l'estimation du vecteur de coefficients β .

La loi de famille exponentielle

La famille exponentielle permet de définir de nombreuses lois connues telles que la loi Normale (pour une variable continue), la loi Binomiale ou la loi de Poisson (pour une variable discrète). Pour Y suivant une loi de famille exponentielle, la fonction de densité s'écrit sous cette forme :

$$f(y; \theta) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} \omega + c(y, \phi) \right\},$$

où θ est le paramètre naturel,

b et c sont des fonctions spécifiques à la loi de Y ,

ϕ est le paramètre de dispersion,

ω est la pondération.

Binomiale sous forme exponentielle : Soit $Y \sim \mathcal{B}(n, \pi)$, donc pour $y \in \{0, \dots, n\}$,

$$\begin{aligned}\mathbb{P}[Y = y] &= \binom{n}{y} \pi^y (1 - \pi)^{n-y} \\ \mathbb{P}[Y = y] &= \exp \left\{ y \ln \left(\frac{\pi}{1-\pi} \right) + n \ln (1 - \pi) + \ln \binom{n}{y} \right\},\end{aligned}$$

et ainsi les paramètres de la famille exponentielle s'écrivent :

$$\begin{aligned}\theta &= \ln \left(\frac{\pi}{1-\pi} \right), \\ \omega &= \phi = 1, \\ b(\theta) &= n \ln \{1 + \exp(\theta)\}, \\ c(y, \phi) &= \ln \binom{n}{y}.\end{aligned}$$

Poisson sous forme exponentielle : Soit $Y \sim \mathcal{P}(\lambda)$, donc pour $y \in \mathbb{N}$,

$$\begin{aligned}\mathbb{P}[Y = y] &= e^{-\lambda} \frac{\lambda^y}{y!} \\ \mathbb{P}[Y = y] &= \exp \{ y \ln(\lambda) - \lambda - \ln(y!) \},\end{aligned}$$

et ainsi les paramètres de la famille exponentielle s'écrivent :

$$\begin{aligned}\theta &= \ln(\lambda), \\ \omega &= \phi = 1, \\ b(\theta) &= \exp(\theta), \\ c(y, \phi) &= -\ln(y!).\end{aligned}$$

Pour la loi multinomiale ordinale, le paramètre de la loi est un vecteur de probabilité. Le procédé est le même que précédemment, cependant il est appliqué au cas multivarié.

Connaissant la loi de $Y|X$, nous avons besoin de définir la fonction de lien.

La fonction de lien

La fonction de lien g connecte le paramètre de moyenne $\mu = \mathbb{E}[Y|X]$ et le prédicteur linéaire $\eta = z^t \beta$,

$$\begin{array}{ccc} g : \mathcal{M} & \rightarrow & \mathbb{R} \\ \mu & \mapsto & \eta \end{array}$$

Nous savons que le prédicteur η est potentiellement compris entre $-\infty$ et ∞ , tandis que le paramètre de moyenne μ est compris dans un espace particulier \mathcal{M} selon la distribution réponse.

Par exemple, pour $Y|X$ suivant une loi de Bernoulli, l'espérance est la probabilité que l'événement arrive et donc $\mathcal{M} = [0, 1]$.

Dans le cas simple de la distribution gaussienne, il n'y a pas de contrainte sur le paramètre de moyenne (μ est compris entre $-\infty$ et $+\infty$) et donc g est la fonction identité.

Pour chaque type de distribution exponentielle, le paramètre naturel θ peut être exprimé en une fonction particulière de paramètre μ . Cette fonction de lien est appelée fonction de lien canonique. Grâce à plusieurs propriétés, tout GLM défini avec le lien canonique est facile à estimer.

Dans le cas de la distribution binomiale de paramètre π , le lien canonique est la fonction logit définie par :

$$g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right)$$

Dans le cas de la distribution de poisson de paramètre λ , le lien canonique est la fonction logarithme :

$$g(\lambda) = \ln(\lambda)$$

Nous venons de voir que les modèles linéaires généralisés utilisant la famille des distributions exponentielle, permettent de définir la fonction de lien.

Finalement, un GLM est caractérisé par : la distribution exponentielle de la variable réponse, la matrice de design Z et la fonction de lien g .

Estimation des paramètres

L'estimation du vecteur de paramètre β est obtenu en maximisant la log-vraisemblance l . Nous savons que dans le cas du modèle linéaire, l'équation $\frac{\partial l}{\partial \beta} = 0$ est résoluble. Mais pour les autres GLM, l'équation n'est pas linéaire en β à cause de la fonction de lien. Ainsi, un algorithme d'optimisation est utilisé pour approximer la solution $\hat{\beta}$. Par exemple, l'algorithme utilisé dans R est celui de Fisher pour l'itération $m+1$:

$$\beta^{[m+1]} = \beta^{[m]} - \left(\mathbb{E} \left[\frac{\partial^2 l}{\partial \beta^t \partial \beta} \right]_{\beta=\beta^{[m]}} \right)^{-1} \left(\frac{\partial l}{\partial \beta} \right)_{\beta=\beta^{[m]}}.$$

L'algorithme converge en un maximum global, s'il existe. Les algorithmes utilisent la fonction de lien en prenant l'inverse de sa dérivé. Cela oblige à ce que la fonction de lien soit un \mathcal{C}^1 -difféomorphisme. C'est bien le cas de la fonction logit et logarithme au vu des valeurs que prennent les variables réponses.

La prédiction

Les GLM sont utilisés pour évaluer l'influence des facteurs sur la variable réponse et pour prédire une valeur sachant les facteurs. Dans le cadre de ce travail, nous souhaitons prédire les valeurs de la variable réponse. Nous cherchons donc à obtenir le paramètre π de la distribution de la variable réponse, afin de pouvoir simuler une réalisation de la variable réponse.

Pour obtenir le paramètre de la distribution réponse, nous savons que l'espérance $\mathbb{E}[Y|X]$ s'exprime en fonction de ce paramètre. Et le calcul de l'espérance est obtenue par les GLM :

$$\mathbb{E}[Y|X] = g^{-1}(Z\hat{\beta}).$$

Par exemple, pour la distribution Binomiale $\mathcal{B}(n, \pi)$:

$$\mathbb{E}[Y|X] = n\pi \quad \text{et donc} \quad \pi = \frac{1}{n}g^{-1}(Z\hat{\beta})$$

Concernant la distribution Multinomiale Ordinale à K modalités, nous obtenons un vecteur de probabilité (π_1, \dots, π_K) à l'aide des GLM.

Pour simuler la loi, nous faisons la somme cumulée du vecteur de probabilité (découpage de l'intervalle $[0, 1]$ en K parties), puis nous simulons une réalisation u de la loi Uniforme. La valeur u se situant dans le $k^{\text{ième}}$ intervalle, nous obtenons une réalisation de la loi multinomiale ordinale.

4.2 Application du modèle au manguier

Lors de la première simulation du manguier sous L-py, nous avons au besoin de quatre informations. Pour la partie végétative, les informations sont : le nombre d'UCs filles que produit une UC mère et leur date d'apparition. Pour la partie reproductive, les informations sont : le nombre d'inflorescence et leur date d'apparition. Le placement des UCs filles et des inflorescences est déterministe et sera abordé plus tard. Dans un premier temps, nous expliquons comment obtenir ces quatre informations à partir des GLM, puis dans un second temps, nous présenterons le modèle de simulation de la croissance.

4.2.1 Les variables du modèle

Dans la section 1.2.2 (État de l'art des analyses de la base Doralice), nous avons vu le processus de croissance végétatif et reproductif (Figure 29).

Le processus biologique de croissance végétative d'une UC mère est le suivant (Figure 29.a) :

- l'UC mère donne-t-elle au moins une UC fille ?
- si oui,
 - ★ à quel date les UCs filles apparaissent ?
 - ★ l'UC mère donne-t-elle au moins une UC fille latérale ?
 - ★ si oui,
 - combien d'UCs filles apparaissent ?

Les questions sont établies pour une UC mère née dans un cycle n .

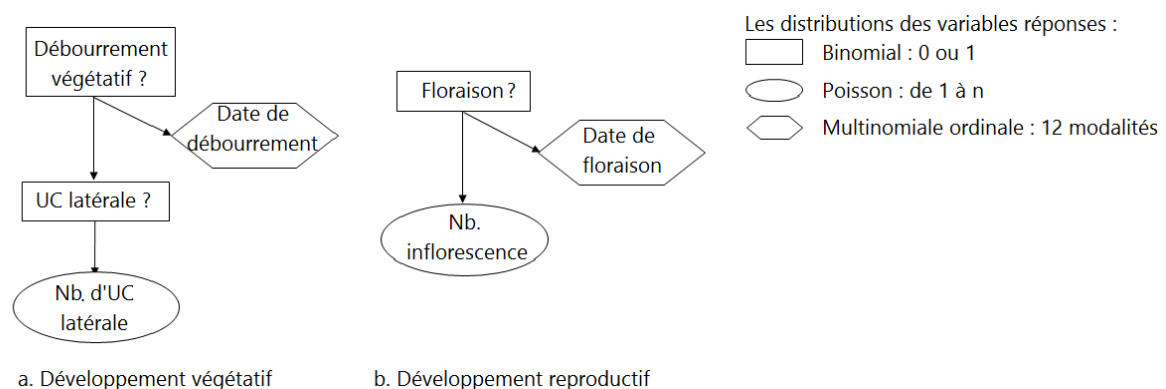


FIGURE 29 – Processus de croissance végétatif et reproductif, (Dambreville, 2012)

Pour la croissance reproductive, les questions sont appliquées seulement pour une UC mère terminale du cycle n . Le processus de questionnement est similaire (Figure 29.b) : L'UC mère donne-t-elle au moins une inflorescence ? Si oui, à quel date et combien d'inflorescences apparaissent ?

4.2.2 Application des GLM

Une variable aléatoire est associée à chacune de ces questions et nous cherchons à répondre aux questions par différents facteurs à l'aide des GLM. Le choix des modèles linéaires généralisés est dû aux distributions des variables réponses (binomiale, poisson et multinomiale ordinaire). Afin de réduire le nombre de modalités de la loi multinomiale ordinaire pour la variable indiquant la date de débourrement, nous avons considéré la différence de temps d'apparition entre l'UC mère et ses UCs filles. Pour le cycle 4, l'écart est de 5 mois maximum.

Le choix des facteurs s'est basé sur les analyses précédentes du stage. Voici les facteurs choisis, avec leurs valeurs prises :

- la date de l'UC mère, [7<8<9<10<11<12<1<2<3<4<5<6] (numéro du mois ordonné selon le cycle)
- la position de l'UC mère, [Apicale, Latérale]
- la charge en fruit, [Chargé, NonChargé]
- la nature de l'ancêtre [Végétatif, Florifère]
- la position de l'ancêtre [Apicale, Latérale]

Le choix de certains facteurs pose problème dans deux cas de modélisation avec les GLM.

Pour passer du cycle de reproduction végétative n au cycle de reproduction végétative $n+1$, ce sont les mêmes questionnements qui reviennent. Le problème est que, lors de la transition des cycles, l'UC mère est confondue avec l'UC ancêtre. Donc le facteur position de l'UC mère est le même que la position de l'ancêtre. Il a fallu distinguer deux cas :

- la croissance au sein du cycle en prenant les UCs des étages 2 et plus
- la croissance de transition où les UCs mères terminales du cycle n donnent des UCs filles du cycle $n+1$

Pour la transition, le facteur "position de l'UC mère" est enlevé, il reste donc : la charge en fruit, la nature de l'ancêtre, la position de l'ancêtre et la date de l'ancêtre.

Dans un second cas, un autre facteur n'est pas pris en compte. En effet, nous souhaitons faire un modèle de simulation qui soit applicable sur un manguier quelconque à la fin du repos végétatif. Les informations que nous avons, sur les MTG mesurés par exemple, sont la position et la nature des UCs, mais nous ne connaissons pas leurs dates d'apparition. Ainsi pour la première croissance de transition (cycle 3 à cycle 4), les seuls facteurs pris en compte sont : la charge en fruit, la nature de l'ancêtre et la position de l'ancêtre.

Nous avons donc établi la croissance séquentielle (relation mère-fille) en deux phases : au sein du cycle et entre les cycles. Mais au niveau des données, il y a une différence entre le cycle 4 et le cycle 5, car les manguiers étudiés étaient encore jeunes et évoluaient d'un cycle à un autre. Nous avons donc établi les deux phases pour chaque cycle.

Ainsi pour chacun de ces 4 cas, nous établissons les GLM sur les variables réponses.

Pour une variable réponse, nous appliquons un GLM à l'aide des fonctions *glm()* et *vglm()* dans R. La fonction *glm()* est utilisée pour estimer le paramètre des lois binomiale et de poisson. La fonction *vglm()* de la librairie **VGAM** est utilisée pour la loi multinomiale ordinaire.

À partir d'un premier modèle pour une variable réponse, nous sélectionnons les facteurs influant la variable par la fonction *step()*. La fonction utilise le critère AIC (Akaike Information Criterion) pour sélectionner les facteurs. Le choix de ce critère est dû au fait que nous souhaitons prédire la variable à l'aide de son GLM. Le critère BIC (Bayesian Information Criterion) pénalise plus les facteurs et est orienté pour l'explication de l'influence des facteurs sur la variable.

4.2.3 La prédiction du paramètre de la distribution réponse

is_loaded	burst_date_mother	position_mother	probas
yes	1	A	0.0398024365420565
no	1	A	0.0563410496322314
yes	2	A	0.0606520440356244
no	2	A	0.0850863448572753
yes	3	A	0.00734038578989795
no	3	A	0.0105384866585414
yes	4	A	2.61565232751709e-08
no	4	A	3.76739210415037e-08
yes	5	A	0.112358797508222
no	5	A	0.15420412862617
yes	6	A	0.999999988342139
no	6	A	0.999999991906096
yes	7	A	0.999999988342139
no	7	A	0.999999991906096
yes	9	A	0.816870595645055
no	9	A	0.865315265704
yes	10	A	0.773908444385843
no	10	A	0.831372112985279
yes	11	A	0.919411271173464
no	11	A	0.942634970876469
yes	12	A	0.695337809063631
no	12	A	0.766752455512396
yes	1	L	0.0201454136032062
no	1	L	0.0287608392460147
yes	2	L	0.0310309001710539
no	2	L	0.0440921509191786
yes	3	L	0.00365421520084364
no	3	L	0.00525480606848342
yes	4	L	1.29731524703719e-08
no	4	L	1.86855691608413e-08
yes	5	L	0.0590732601894323
no	5	L	0.0829276817269365
yes	6	L	0.999999976495372
no	6	L	0.999999983681037
yes	7	L	0.999999976495372
no	7	L	0.999999983681037
yes	9	L	0.688704797176301
no	9	L	0.761140156232017
yes	10	L	0.629319120842674
no	10	L	0.709749232167779
yes	11	L	0.849815935682704
no	11	L	0.890711328401312
yes	12	L	0.530954758722036
no	12	L	0.619834799097705
yes	8	A	0
no	8	A	0
yes	8	L	0
no	8	L	0

FIGURE 30 – Tableau des facteurs et de la probabilité de la variable de débourrement au cycle 4

Maintenant que nous avons estimé un modèle pour chaque variable réponse, nous souhaitons avoir le paramètre de sa distribution selon des facteurs donnés. La fonction *predict()* de R prend, en entrée, les valeurs connues des facteurs et fournit, en sortie, le paramètre de la distribution réponse.

Afin de faciliter la simulation de croissance sous L-py, nous établissons un tableau qui contient le produit cartésien des valeurs prises par les facteurs (Figure 30). Il y a 192 configurations possibles pour les valeurs prises par les facteurs. Pour chaque cas, nous récupérons le paramètre de la distribution à l'aide de la fonction *predict()*. Nous ajoutons une colonne qui contient le paramètre associé à une configuration de facteurs. Les configurations ne sont pas toutes observées dans la base de données. Dans ce cas, il est impossible de prédire la valeur du paramètre et nous lui attribuons la valeur 0, car c'est un phénomène improbable.

4.2.4 Simulation du nombre d'UCs filles et leur date d'apparition

Pour obtenir le nombre de filles et leur date d'apparition, nous suivons les schémas de la figure 29.

Pour une UC mère, plusieurs facteurs sont connus : sa position (A ou L), son mois d'apparition (septembre à mai), la nature de son ancêtre (V ou F) et la position de son ancêtre (A ou L).

Grâce à ces informations, le procédé de simulation du nombre d'UCs filles et de leur date d'apparition se met en place :

- * La question est de savoir si l'UC mère débouffe. A partir des facteurs connus, nous estimons le paramètre π_B de la loi Binomiale $\mathcal{B}(1, \pi_B)$ puis, nous obtenons une réalisation de cette loi.
- * Si la réalisation vaut 1 (oui, il y a débouffement), il y a au moins une UC fille
 - La question est de savoir s'il y a des UCs latérales. Le procédé est le même : récupération du paramètre π_L et simulation de la loi Binomiale $\mathcal{B}(1, \pi_L)$.
 - Si la réalisation vaut 1 (oui, il y a au moins une UC latérale),
 - ★ La question est de savoir combien d'UCs latérales produit l'UC mère. Nous récupérons le paramètre λ associé aux facteurs connus et nous obtenons une réalisation n_i de la loi de Poisson. Nous obtenons ainsi le nombre d'UCs filles : $1 + n_i$.
 - Sinon, (non, il n'y a pas d'UC latérale),
 - ★ Le nombre d'UCs filles est de 1.
 - La question est de savoir à quelle date les UCs filles apparaissent. Nous récupérons le vecteur de probabilité associé aux facteurs connus puis, nous simulons une réalisation de la loi multinomiale ordinaire. Nous obtenons ainsi la différence d'apparition entre l'UC mère et ses UCs filles, à l'échelle du mois.
- * Sinon, (non l'UC mère n'a pas débouffé dans ce cycle), l'UC mère se trouve en position terminale, nous nous situons dans la phase de transition de cycle.
 - La question est de savoir si elle fleurit. Nous prenons la probabilité π_F associé aux facteurs et nous simulons une loi Binomiale $\mathcal{B}(1, \pi_F)$.
 - ★ Si la réalisation vaut 1, (l'UC fleurit),
 - La question est de savoir combien d'inflorescences apparaissent. Nous simulons une loi de Poisson à partir du paramètre estimé par les GLM. Nous obtenons le nombre d'inflorescences.
 - La question est de savoir à quelle date les inflorescences apparaissent. Le vecteur de probabilité est récupéré et nous simulons une réalisation de la loi multinomiale ordinaire. Nous obtenons ainsi une semaine d'apparition des inflorescences, entre août et octobre.
 - ★ Sinon, (l'UC ne fleurit pas), il ne se passe rien
 - ★ L'UC mère devient l'UC ancêtre pour le cycle suivant. Nous réactualisons les facteurs de la position de l'ancêtre et de la nature de l'ancêtre. Si elle a produit des inflorescences, sa nature est florifère, sinon elle est végétative.
 - La question est de savoir si l'UC débouffe au cycle suivant. Nous récupérons le paramètre associé aux facteurs et nous simulons une loi binomiale.

- ★ Si la réalisation vaut 1, (l'UC débourre au cycle suivant). Nous déterminons la date d'apparition et le nombre d'UCs filles de la même façon que précédemment. Il y a une nuance pour le nombre d'UCs filles, car cela dépend de la nature de l'UC mère. Si la nature est florifère, le nombre d'UCs filles est le nombre de latérales. Sinon, c'est le nombre de latérales +1.
- ★ Sinon, (l'UC ne débourre pas). Nous considérons que l'UC meurt au bout de deux cycles, et nous calculons sa date de mort en ajoutant deux ans à sa date d'apparition. L'UC n'est pas supprimé pour la validation de la simulation de croissance.

A l'aide de ce cheminement, nous obtenons le nombre d'UCs filles, leur date d'apparition si elle existe, le nombre d'inflorescences si c'est le cas et la date d'apparition des inflorescences s'il y en a.

4.2.5 Modélisation de la croissance dans L-py

Lors de la modélisation de la croissance sous L-py, nous avons besoin de déterminer quelle entité il faut placer en position apicale. La priorité de la position apicale s'applique aux inflorescences. Nous cherchons donc en premier lieu si l'UC mère donne des inflorescences. Si c'est le cas, une inflorescence est placée en position apicale et les autres inflorescences et UCs filles sont réparties uniformément autour de l'inflorescence apicale. Sinon, une UC fille est placée en position apicale et les autres UCs sont placées uniformément autour de l'UC fille apicale. La règle de production s'écrit donc de cette façon :

```

1 A(burst_date, is_loaded, position_mother, position_ancestor,
2   nature_ancestor):
3   if burst_date <= current_date :
4       burst_date = current_date
5       nproduce UC(burst_date,nature)
6       nb_children, date_children, nb_inflo, date_inflo =
7           get_simulation(burst_date, position_mother, is_loaded,
8                           nature_ancestor, position_ancestor)
9       nb_total = nb_children + nb_inflo
10      if nb_total > 1:
11          angle_between_children = 360./nb_total
12      if nb_inflo > 0 :
13          if nb_inflo > 1:
14              for inflo in xrange(0,nb_inflo-1):
15                  nproduce /(angle_between_children) [&(60)
16                      Inflorescence(date_inflo)]
17          if nb_children > 0:
18              for uc in xrange(0,nb_children):
19                  nproduce /(angle_between_children) [&(60) A(date_children,
20                      "L", is_loaded, "F", position_mother)]
21          produce Inflorescence(date_inflo)
22      else :
23          if nb_children > 0:
24              if nb_children > 1:
25                  for uc in xrange(0,nb_children-1):
26                      nproduce /(angle_between_children) [&(60)
27                          A(date_children, "L", is_loaded, "V",
28                              position_ancestor ) ]

```



```

29         produce A(date_children, "A", is_loaded, "V",
30                   position_ancestor)

```

Les lignes 1 à 9 récupère les informations nécessaires pour placer les entités. La fonction *get_simulation()* prend en entrée les facteurs qui permettent de prédire les valeurs des variables réponses vues précédemment. Elle renvoie le nombre d'UCs filles, leur date d'apparition, le nombre d'inflorescences et la date d'apparition des inflorescences. De la ligne 12 à la ligne 21, nous sommes dans le cas de la transition de cycle. Tandis que de la ligne 22 à 30, nous sommes dans le cycle de production végétative.

Les règles d'interprétation des modules UC et Inflorescence sont les mêmes que pour le modèle simple de croissance vu à la section 3.2.

Un premier résultat visuel :



FIGURE 31 – La croissance du manguiier pendant les deux cycles de production 4 et 5

Le premier arbre (en haut, à gauche) de la figure 31 est la reproduction de l'arbre B12 en 2003 juste avant le prélèvement des données.

Ensuite, afin de valider le modèle de croissance, 1 000 simulations de ce modèle sont faites. Quelques résultats sont à la section 6, (Résultats préliminaires).

5 Modélisation basée sur les modèles graphiques

A présent, nous ne supposons plus que les UC filles apparaissent en même temps sur une même UC mère. Il y a plusieurs vagues de croissance qu'il faut prendre en compte et cela complique le schéma de processus de croissance (questionnement d'une seconde vague, d'une troisième vague,...). L'utilisation des modèles graphiques permet de simplifier ce schéma.

Dans ce chapitre, nous abordons de façon générale les modèles graphiques, puis nous expliquons le processus de réflexion qui a amené à la simulation basée sur les modèles graphiques.

5.1 Le modèle statistique

Soient X_1, \dots, X_n des variables aléatoires telles que certaines variables soient indépendantes conditionnellement à d'autres variables. Dans des cas complexes, il est intéressant de représenter les liens d'indépendance par un graphe.

Les modèles graphiques

Un **modèle graphique** est un modèle probabiliste composé d'un graphe et d'une loi de probabilité associée au graphe. Une propriété de factorisation de la distribution fait le lien entre le graphe et la distribution.

Formellement, un graphe G s'écrit $G = (\mathcal{U}, \mathcal{E})$ où $\mathcal{U} \subset \mathbb{N}$ est un ensemble de sommets et $\mathcal{E} \subset \mathcal{U} \times \mathcal{U}$ est un ensemble d'arêtes.

Nous nous intéressons à un modèle graphique particulier : les graphes acycliques orientés (DAG : Directed Acyclic Graph). Il est dit que i est le **parent** de j si le couple (i, j) appartient à l'ensemble \mathcal{E} des arêtes de G . Les DAG impliquent une propriété sur les arêtes : Soient i et $j \in \mathcal{U}$, si $(i, j) \in \mathcal{E}$, alors $(j, i) \notin \mathcal{E}$. L'ensemble des parents de $X_i, \forall i \in \mathcal{U}$ est noté $pa_i = \{j \in \mathcal{U}, (i, j) \in \mathcal{E}\}$.

Lorsque la structure du modèle est un DAG, la propriété de factorisation s'écrit de la façon suivante : Soient X_1, \dots, X_n les variables aléatoires et x_1, \dots, x_n leurs réalisations, alors la probabilité jointe satisfait

$$\mathbb{P}[X = x] = \prod_{i=1}^n \mathbb{P}[X_i = x_i | (X_j = x_j)_{j \in pa_i}]$$

où X est le vecteur aléatoire (X_1, \dots, X_n) et

pa_i est l'ensemble des parents de X_i .

En d'autre termes, la distribution jointe se factorise en un produit de distributions marginales ($\text{card}(pa_i) = 0$), et de distributions conditionnelles ($\text{card}(pa_i) > 0$). Cette propriété de factorisation induit des indépendances.

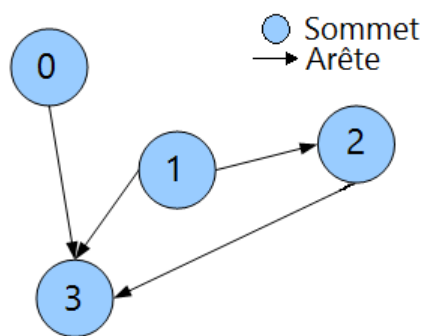


FIGURE 32 – Exemple de modèle graphique : graphe acyclique orienté (DAG), les arêtes indiquent les liens.

Par exemple, le modèle graphique de la figure 32, se compose des variables aléatoires X_0, X_1, X_2, X_3 . Chaque variable aléatoire est représentée par un cercle et les arêtes sont représentées par des flèches. Les sommets 0 et 1 sont indépendants marginalement mais dépendants conditionnellement au sommet 3. La probabilité jointe se factorise :

$$\mathbb{P}[X = x] = \mathbb{P}[X_0 = x_0] \mathbb{P}[X_1 = x_1] \mathbb{P}[X_2 = x_2 | X_1 = x_1] \\ \mathbb{P}[X_3 = x_3 | (X_0, X_1, X_2) = (x_0, x_1, x_2)]$$

Pour obtenir une réalisation de la loi de probabilité d'un graphe, il faut d'abord obtenir la réalisation des sommets i où $\text{card}(pa_i) = 0$ (sommets indépendant marginalement), puis la réalisation des sommets j où $\text{card}(pa_j) = 1$, et ainsi de suite. Un ordre topologique se met en place.

Par exemple, pour le graphe de la figure 32, pour avoir une réalisation du sommet 2, il faut d'abord avoir une réalisation du sommet 1.

Lors de la description des modèles graphiques, nous supposons que le graphe est connu. Dans le prochain paragraphe, nous expliquons une méthode de sélection de modèle graphiques lorsque nous sommes confrontés à des données.

Le choix d'un graphe optimal

Il n'est pas facile de trouver un graphe correspondant aux données observées, un compromis se met en place. Il faut trouver les liens de dépendance entre les variables qui représentent le plus les phénomènes observés, tout en cherchant à simplifier les relations de dépendance.

Pour rechercher un graphe optimal, la procédure utilisée dans le cadre de la modélisation est décrite comme suit :

Etape 1 : choisir un graphe initial $G^{(0)}$ quelconque

Etape 2 : appliquer la fonction *search*, (opération sur les arêtes, figure 32)

Etape 3 : appliquer la fonction *score* sur les différents graphes

Etape 4 :

- **Si** un des scores des différents graphes proposés est meilleur que le score actuel : choisir le graphe ayant le meilleur score et aller à l'étape 2
- **Sinon**, prendre le graphe ayant le meilleur score et s'arrêter.

Explicitons la fonction *search* de l'étape 2 : à partir d'un graphe $G^{(t)}$, la fonction renvoie une collection de graphes $(G_i^{(t)})$. Un graphe $G_i^{(t)}$ est obtenu, à partir du graphe $G^{(t)}$, en effectuant une seule opération sur les arêtes. Une opération sur les arêtes est la suppression d'une arête, l'ajout d'une arête ou le changement de sens d'une arête. Une condition est importante, il faut que le graphe reste acyclique. Un exemple sur la figure 33, montre les 6 possibilités de graphes avec seulement 3 sommets. La fonction *search* appliqué au graphe $G^{(t)}$ (graphe dont les sommets sont vert), donne les différents graphes.

La fonction *score* de l'étape 3 calcule un score pour chaque graphe. Ici, le score est calculé à l'aide du critère BIC. Le choix de ce critère par rapport au critère AIC est que cela permet d'avoir moins d'arêtes sur le graphe final et donc d'avoir moins de dépendances. L'interprétation en sera facilitée pour les biologistes et permettra de discuter sur

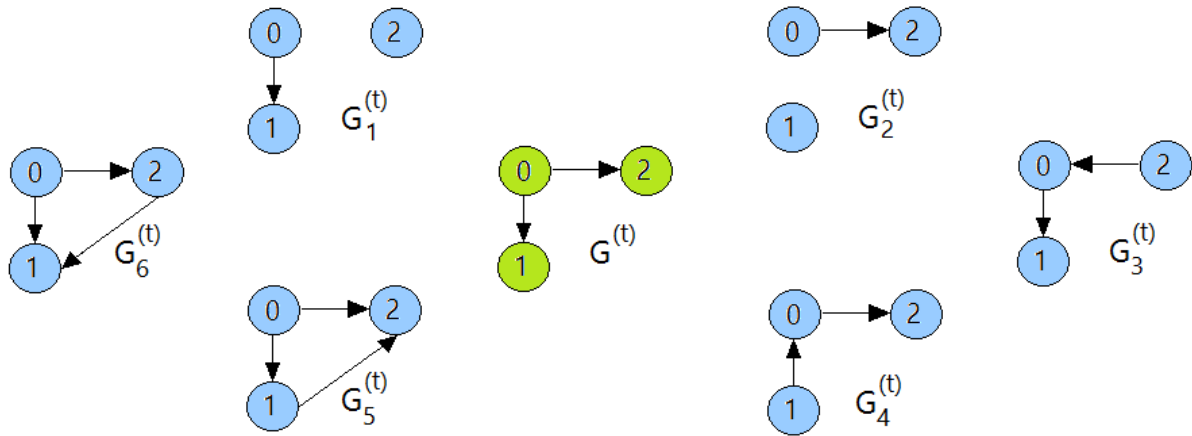


FIGURE 33 – Exemple de résultats de la fonction *search* sur le graphe $G^{(t)}$, (sommets vert)

le graphe (cohérence avec la connaissance biologique).

Étant donné que la méthode dépend du point de départ, l'optimum de la méthode est local. Pour contrer ce problème, la méthode est utilisée plusieurs fois avec un départ aléatoire. D'autres méthodes plus optimales font l'objet de recherches.

5.2 Application du modèle au manguier

Les états des UCs : sommets du graphe

L'utilisation des modèles graphiques est faite de façon à connaître le nombre d'UCs filles que produit une UC mère. Une réalisation de la loi d'un graphe donne le nombre d'enfants d'une UC possédant certaines caractéristiques. C'est donc une loi de comptage qui est attribuée pour chaque sommet.

La nature et la temporalité d'une UC ont été choisies pour la caractériser. La nature de l'UC peut être végétative (elle donne uniquement des UCs) ou florifère (elle donne au moins une inflorescence). La position de l'inflorescence (apicale ou latérale) est aussi prise en compte. C'est pourquoi, une distinction est faite entre les UCs qui donnent une seule inflorescence (qui est en position apicale) et les UC qui donne plusieurs inflorescences (une inflorescence en position apicale et les autres en position latérale).

La temporalité d'apparition de l'UC se distingue en deux phases :

1. distinction entre les cycles de production :

Immediate : au sein du cycle, par rapport à l'UC mère

Delayed : au cycle suivant, par rapport à l'UC mère

2. distinction au sein d'un cycle de production, trois périodes :

Early : de juillet à octobre de l'année $n - 1$

Intermediate : de novembre de l'année $n - 1$ à février de l'année n

Late : de mars à mai de l'année n

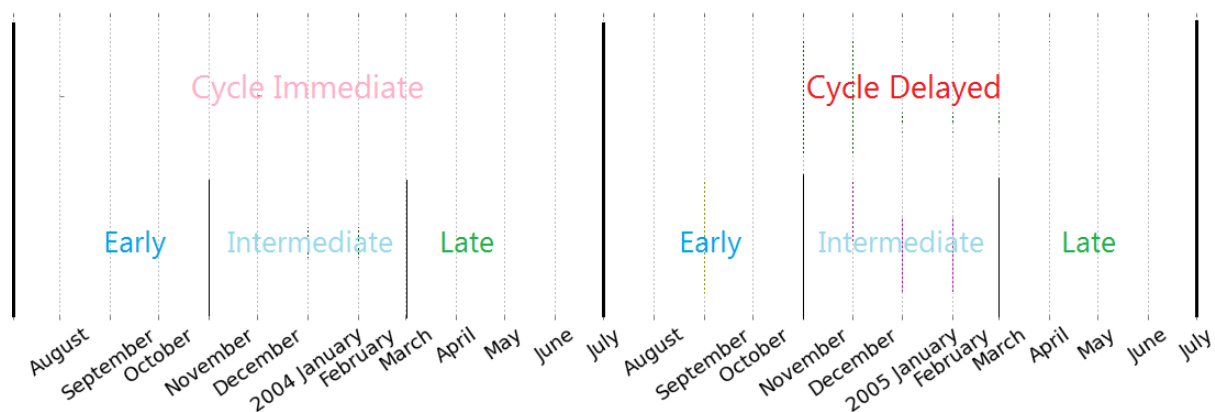


FIGURE 34 – Frise chronologique pour une UC mère née dans le cycle Immediate

Pour le premier point, si l'UC apparait durant le même cycle que son UC mère, elle est dite **Immediate**. Si l'UC fille apparait au cycle suivant, elle est dite **Delayed**.

Pour le second point, les trois périodes ont été délimitées pour prendre en compte les éventuelles compétitions dues à la superposition des cycles de production. La première période, **Early**, correspond à une période de possible compétition entre la floraison et la croissance végétative. La seconde période, **Intermediate**, correspond à une période de possible compétition entre la fructification et la croissance végétative. La dernière période, **Late**, concerne seulement la croissance végétative.

Une dernière temporalité **Unknown** a été introduite, car les dates d'apparition des UCs du cycle 3 sont inconnues. Donc nous distinguons les UCs du cycle 3, où la temporalité est inconnue, par rapport aux UCs des cycles suivants. C'est pourquoi nous créons deux états : UV et UF.

Le nom des variables est une abréviation de l'anglais. Elles sont dans le tableau suivant :

Variable	Nature de l'UC	Temporalité du cycle	Temporalité dans un cycle
UV	vegetative	unknown	unknown
IIV	vegetative	immediate	intermediate
ILV	vegetative	immediate	late
DEV	vegetative	one-cycle delayed	early
DIV	vegetative	one-cycle delayed	intermediate
DLV	vegetative	one-cycle delayed	late
UF	flower	unknown	unknown
IIT	Terminal flower	immediate	intermediate
ILT	Terminal flower	immédiate	late
DIT	Terminal flower	one-cycle delayed	intermediate
DLT	Terminal flower	one-cycle delayed	late
IIL	Lateral flower	immédiate	intermediate
DIL	Lateral flower	one-cycle delayed	intermediate

Cela nous donne 13 états (sommets). Toutes les combinaisons ne sont pas prises en compte, car soit le cas n'a pas été observé ou, soit c'est un cas trop rare pour être pris en compte (par exemple, les cas IEV, DET et LIL). Dans le dernier cas, ces états ont été associés à l'état le plus proche.

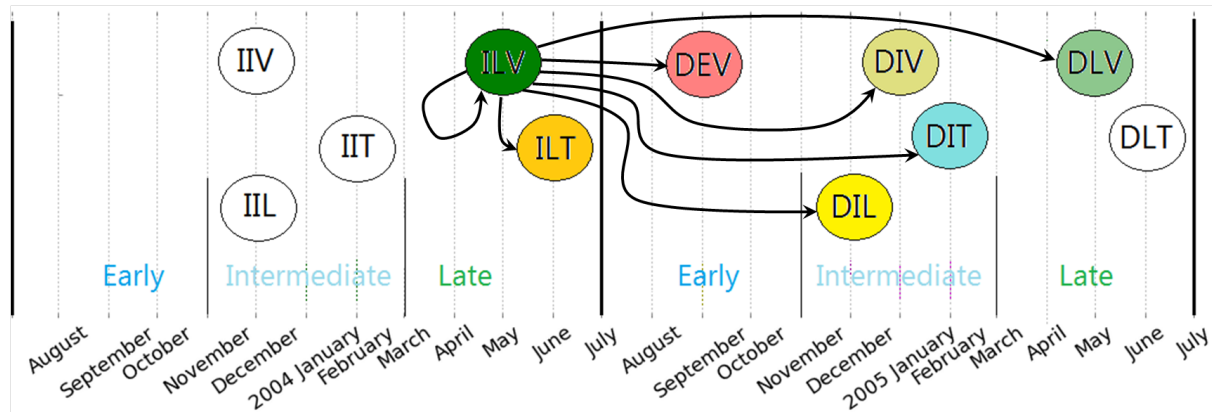


FIGURE 35 – Pour une UC dans l'état ILV, les UCs filles possible qu'elle donne, sont en couleur

Un graphe est associé à chaque état. Plus précisément, le graphe associé, à une UC dans un certain état, indique toutes les différentes UCs filles qu'elle peut avoir. Ainsi, la variable associée à un état suit une loi de comptage. Si c'est une distribution marginale, les lois possibles sont la loi de Poisson, la loi Binomiale ou la loi Binomiale Négative. Le choix de la distribution est fait à l'aide du maximum de vraisemblance. Sinon, c'est une distribution conditionnelle et les lois possibles sont la régression de Poisson, la régression Binomiale ou la régression Binomiale Négative. La régression se fait à l'aide des GLM.

Soit une UC mère dont son état est ILV (Immediate Late Vegetative), le graphe de la figure 36, indique toutes les UCs filles quelle peut donner et leurs distributions associées.

Les états du graphe qui sont positionnés à gauche ont tous une distribution Binomiale de paramètres $n = 2$ et $\pi = 0$. Cela indique que l'UC mère ne produira pas d'UCs filles de ces états. Par exemple, l'UC mère apparue dans la période Late, ne produira pas d'UC filles, au même cycle, à la période Intermediate. Les états du graphe qui sont positionnés à droite ont des distributions marginales. Par exemple, l'UC mère de la période Late peut donner des UCs filles dans la même période selon une distribution de Poisson de paramètre $\lambda = 0.015$. Au centre du graphe, nous avons l'influence de certaines UCs filles sur ses sœurs. Les états estimés par une régression sont indiqués en donnant la fonction de lien utilisée et les coefficients du paramètre de régression. Par exemple, si l'UC mère produit une UC fille végétative au cycle suivant, à la période Late, l'UC mère aura une probabilité très faible de donner une UC fille florifère à la période Intermediate.

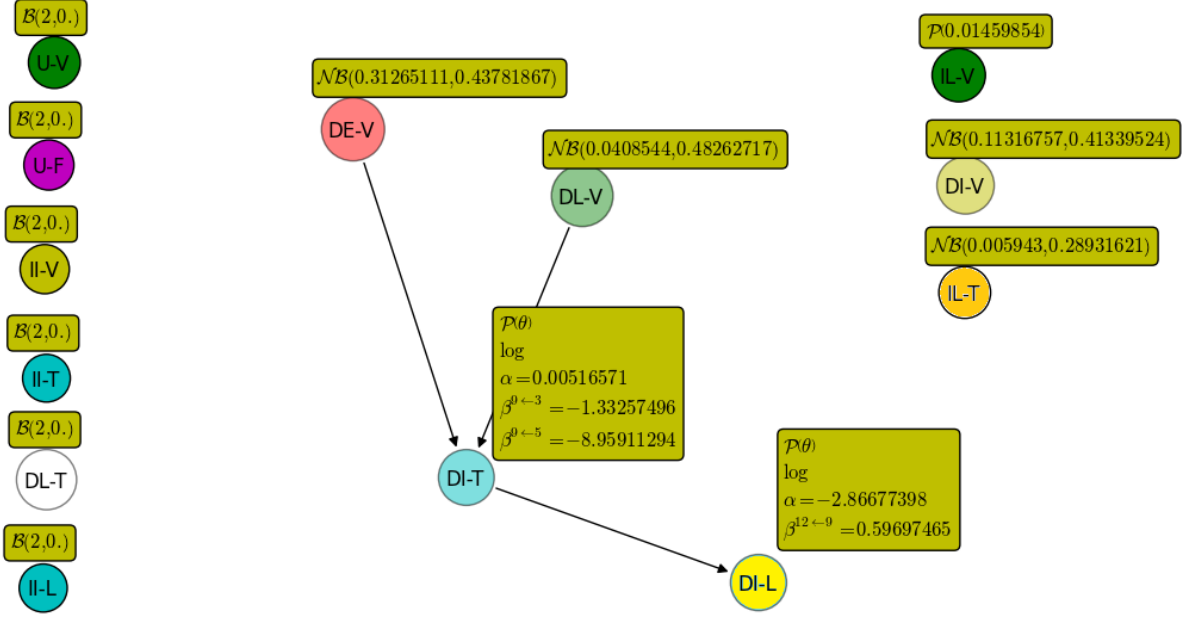


FIGURE 36 – Graphe d'une UC mère ILV

Les mélanges finis

La notion de modèle de mélange a été introduite dans l'utilisation des graphes, car certaines variables sont caractérisées par des distributions mélangées. En effet, le nombre d'enfants par unité de croissance ne suit pas exactement une loi de Poisson. Nous observons une abondance de 0 enfant. Il y aurait donc un mélange de deux distributions connues : la loi de Dirac en 0 et la loi de Poisson.

Une variable X suit un modèle de mélange lorsque sa distribution provient d'une association de $k \in \mathbb{N}^*$ distributions connues. Sa densité s'écrit sous cette forme :

$$f(x) = \sum_{k=1}^K \pi_k f_k(x) \quad \text{et} \quad \sum_{k=1}^K \pi_k = 1$$

où f_k est la densité de la $k^{ième}$ loi et

π_k est la probabilité de tomber dans la $k^{ième}$ distribution.

f est bien une densité de probabilité car

$$\begin{aligned} \int_{\Omega} f &= \int_{\Omega} \sum_{k=1}^K \pi_k f_k \\ &= \sum_{k=1}^K \pi_k \int_{\Omega} f_k \\ &= \sum_{k=1}^K \pi_k \end{aligned}$$

C'est pourquoi la somme des poids π_k doit être égale à 1.

Prenons l'exemple de l'association d'une loi de Dirac en 0 et d'une loi de Poisson de paramètre λ . Prenons arbitrairement le poids de la loi de Poisson égale à 2/3. Soit l'état 1 où X_1 suit la loi de Dirac : $\mathbb{P}_1[X_1 = 0] = 1$ et $\mathbb{P}_1[X_1 \neq 0] = 0$. Soit l'état 2 où X_2 suit

la loi de Poisson de paramètre λ : $\mathbb{P}_2[X_2 = x] = \frac{\lambda^x}{x!}e^{-\lambda}$. La distribution est donc

$$f(x) = \pi_1 \mathbb{P}_1[X_1 = x] + \pi_2 \mathbb{P}_2[X_2 = x]$$

La simulation d'un modèle de mélange se fait en deux étapes. La première consiste à savoir dans quel état nous sommes et la seconde étape simule la loi correspondant à l'état obtenu. Pour obtenir un état parmi K possibilités, nous simulons une loi multinomiale à K modalités et de vecteur de probabilité (π_1, \dots, π_K) . Soit $k \in \{1, \dots, K\}$ le résultat, nous simulons ensuite la k^{ime} loi. La valeur obtenue est une réalisation du modèle de mélange.

Sachant que la charge en fruits de l'arbre influe sur le comportement de croissance, la distinction entre les arbres chargés et les non chargés a donné deux possibilités pour simuler la croissance. De plus, certains états suivent une distribution de mélange. Pour comparer l'effet de l'utilisation des modèles mélangés, nous avons la possibilité d'avoir les graphiques avec et sans les modèles mélangés. Cela implique au total, 4 possibilités de simulation (arbres chargés/non chargés et modèles de mélange ou non) à partir des différents modèles graphiques.

5.3 Les procédés de simulation

Dans cette section, nous faisons un état des connaissances et des informations manquantes pour la simulation, puis nous décrivons la génération du modèle de croissance.

Ce qui est acquis et ce qu'il manque

En utilisant la fonction *simulate()* du module statistics, les états de chaque filles, nous permettent de savoir dans quelle période elles apparaissent et si elles donnent des inflorescences. Ceci est décrit par la définition des états.

Plusieurs informations sont manquantes pour pouvoir simuler la croissance, mon travail a été de les récupérer à partir du MTG des données observées. Les informations manquantes sont :

- la date d'apparition de l'UC fille à l'échelle du mois
- la détermination de la position des UCs filles (quelle UC sera en position apicale?)
- la détermination du nombre d'inflorescences latérales
- la détermination de la date de floraison

Acquisition des informations manquantes

Date de débourrement des Ucs à l'échelle du mois :

Une première possibilité est d'effectuer un tirage uniforme sur la période. Ce procédé est simple mais ne représente pas bien les données.

La deuxième possibilité est d'effectuer un tirage gaussien sur la période. La moyenne de la gaussienne est le milieu de la période et la variance est égale à deux mois. Mais

le problème de cette distribution est quelle n'est pas bornée. Il est possible d'avoir un écart de temps d'une semaine entre l'UC mère et les UCs filles. Pour contrer ce problème, un écart minimal de 6 semaines a été instauré entre l'apparition de l'UC mère et de ses UCs filles. Il y a aussi un autre problème, il arrive d'observer qu'une UC mère apparaisse au début d'une période et qu'elle donne des UCs filles à la fin de la même période. Or pour la simulation, lorsqu'une UC mère apparaît à la fin de la période et doit donner des filles dans cette même période, un tirage gaussien avec pour moyenne un décalage de 8 semaines (± 2 semaines) entre les dates de l'UC mère et de l'UC fille a été mis en place. Les résultats de cette méthode n'ont pas été concluants.

La troisième possibilité est de prendre la distribution empirique des données par période.

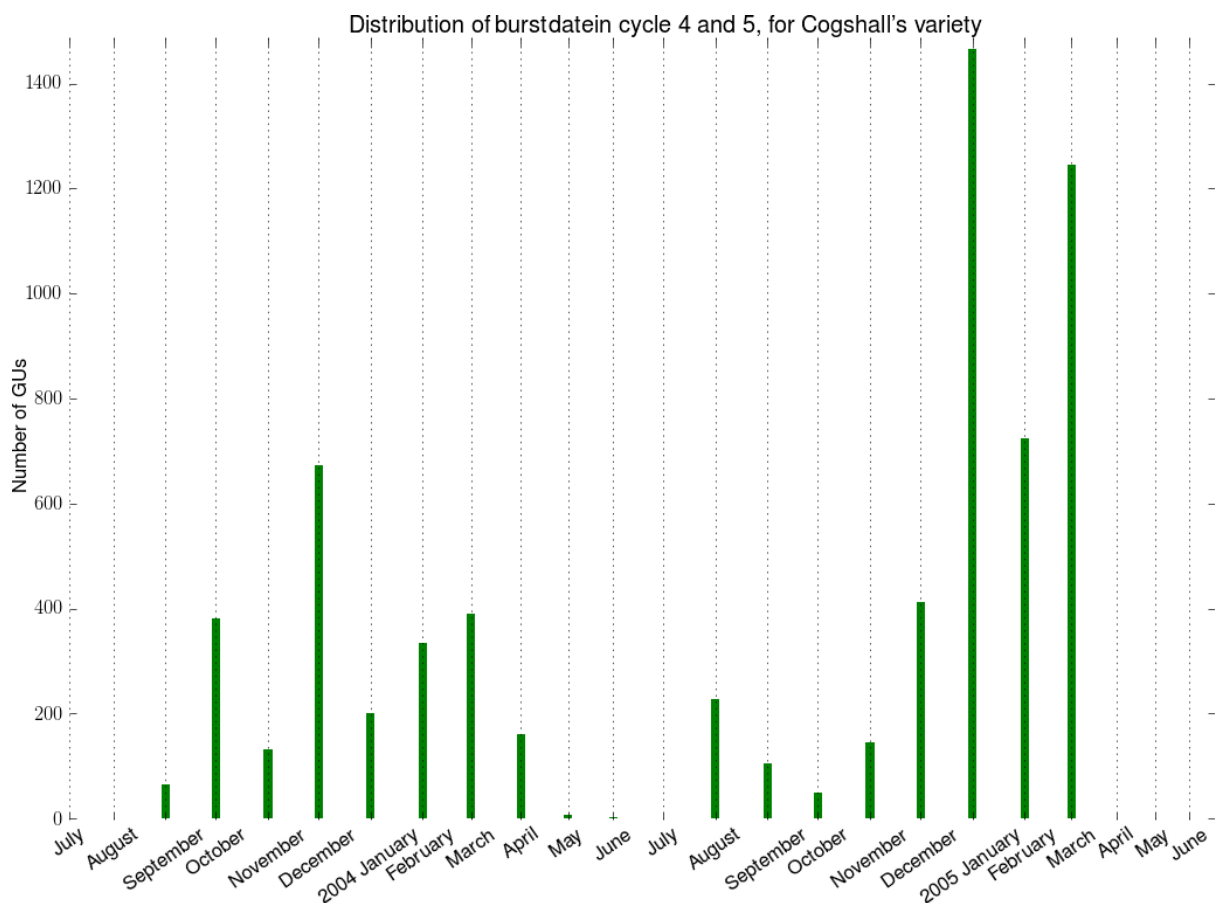


FIGURE 37 – Distribution des dates d'apparition des UCs, au cycle 4 et 5, pour la variété Cogshall

La dernière possibilité est d'ajuster la distribution d'apparition des UCs dans une période. Dans certaines périodes, ces distributions sont des distributions mélangées. Par exemple, sur la figure 37, il y a différentes gaussiennes qui se superposent.

Cette dernière possibilité sera faite prochainement.

Détermination de l'UC fille en position apicale :

Du fait de la dominance apicale, nous savons que l'UC fille qui apparait en premier est en position apicale sur une UC n'ayant pas fleuri. Nous trions donc la liste des états en fonction de la temporalité (du plus récent au plus éloigné). De plus, les UCs en position apicale ont une probabilité plus élevée de fleurir et leur nombre d'inflorescences ou d'UC fille est plus élevé par rapport à des UCs en position latérale (Normand et al., 2009 ; Dambreville et al., 2013a). L'ordre d'apparition se fait donc en premier pour les UCs donnant plusieurs inflorescences, puis pour les UCs donnant une inflorescence et enfin pour les UCs végétatives. Voici l'ordre de priorité pour le positionnement apicale :

IIL,IIT,IIV,ILT,ILV,DEV,DIL,DIT,DIV,DLT,DLV

Détermination du nombre d'inflorescences latérales :

En visualisant la distribution du nombre d'inflorescences latérales par UC (Figure 38), nous avons testé si la variable du nombre d'inflorescences latérales suit une loi de Poisson. La loi de Poisson commençant à 0, nous avons effectué un décalage de -1 sur les données. La première étape de test a été de comparer la moyenne et la variance pour chaque cycle. Puisque la moyenne et la variance ne sont pas proches, nous avons testé si la variable suit une loi de quasi-Poisson à l'aide de la fonction *glm()* dans R. La variable suit bien une loi de quasi-Poisson. De plus les paramètres des deux cycles sont très proches, c'est pourquoi, nous ne différencions pas les deux cycles. L'estimation du paramètre λ se fait en calculant la moyenne. La valeur obtenue est de 1,2. A cela, il faut ajouter +1 pour obtenir le nombre moyen d'inflorescences latérales.

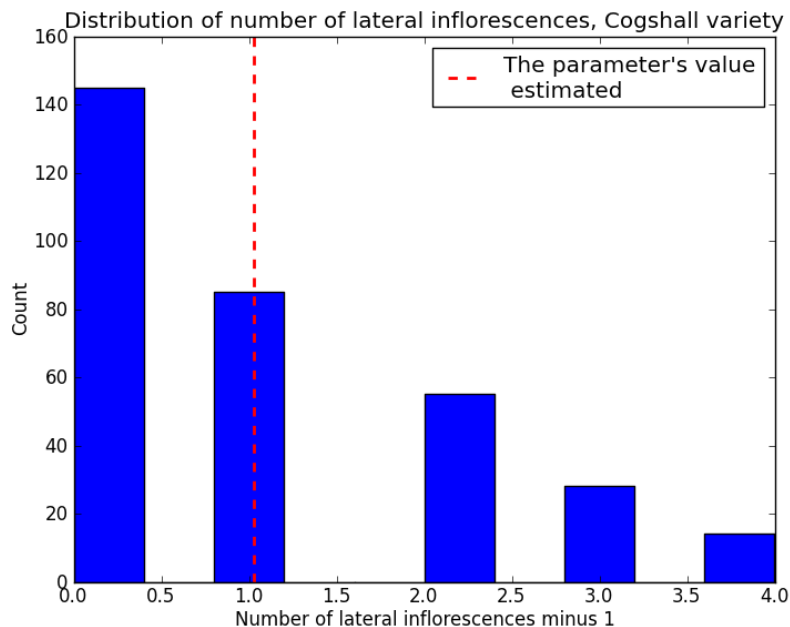


FIGURE 38 – Distribution du nombre d'inflorescences latérales pour la variété Cogshall

Détermination de la date de floraison

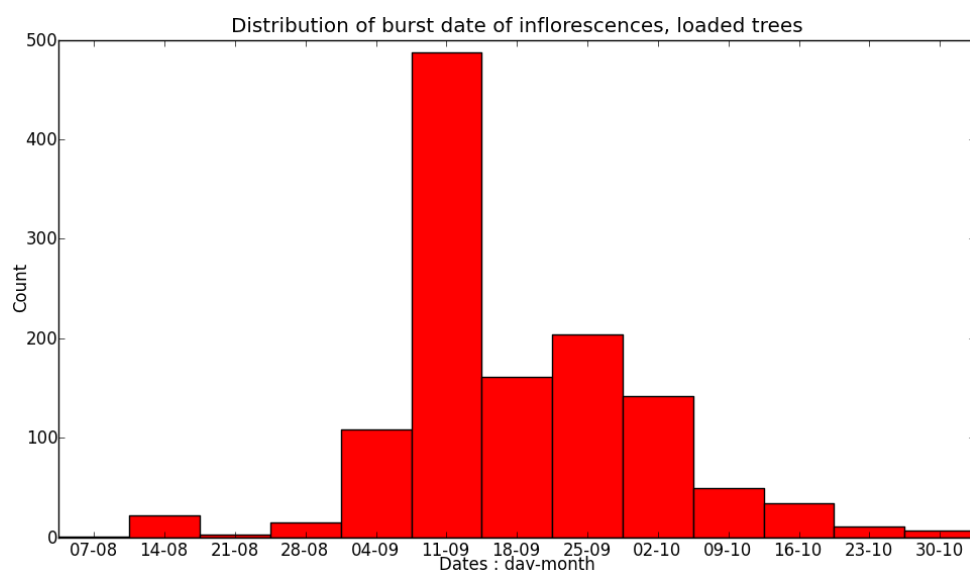


FIGURE 39 – Distribution des dates d'apparition des inflorescences, pour les arbres chargés

Les dates de floraison s'étalent sur trois mois, d'août à octobre. Ayant la date de floraison à l'échelle du jour, nous avons pu rassembler les dates d'une même semaine et obtenir les distributions sur la Figure . Puisque la charge de l'arbre a un impact sur la date de floraison, nous avons fait une distinction entre les arbres chargés et les non chargés. Dans la simulation, nous déterminons la date de floraison à partir de la distribution empirique.

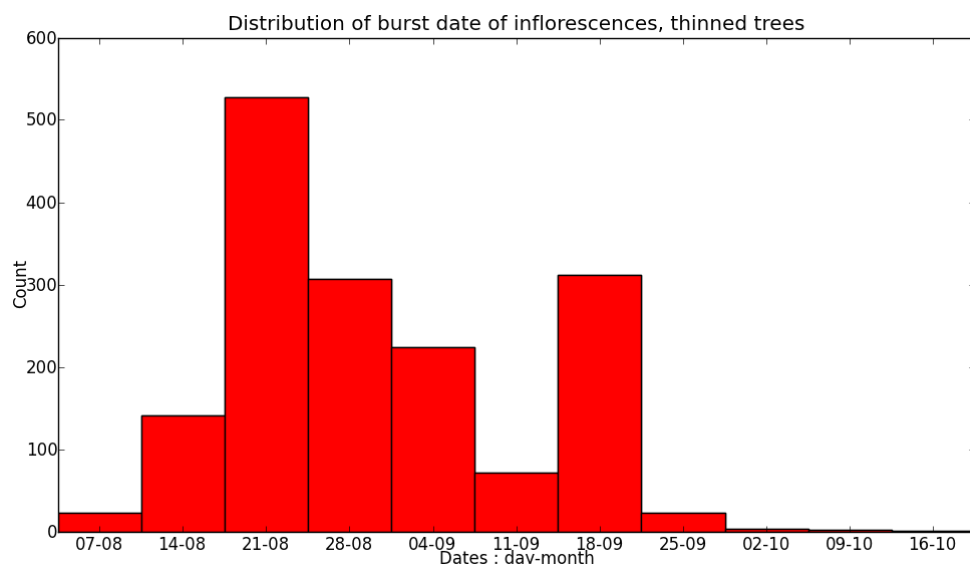


FIGURE 40 – Distribution des dates d'apparition des inflorescences, pour les arbres non chargés

La simulation

Le principe de la simulation est le même pour les précédentes simulations de croissance. Nous plaçons les entités en fonction de la nature de l'UC mère. Si elle est florifère, toutes ces UCs filles sont en position latérale. Le résultat visuel est semblable au résultat basé sur les modèles linéaires généralisés.

Afin de vérifier si la simulation de croissance est semblable aux données du terrain, nous effectuons 1 000 simulations.

6 Résultats préliminaires

Lors de l'estimation des modèles linéaires généralisés, nous avons estimé des premiers modèles comprenant tous les facteurs. Nous nommons ces modèles "GLM complets". Puis à l'aide du critère AIC, des modèles plus parcimonieux ont été estimés où seuls les facteurs influant ont été sélectionnés. Ces modèles sont dit "GLM sélectionnés" (sélection des facteurs influant). Les GLM ont été estimés sur la base de données des cinq arbres de la variété Cogshall. Les modèles graphiques ont été estimés sur les 3 arbres chargés de la variété Cogshall.

Pour valider nos simulations, nous avons mis en place la méthode présentée dans la section suivante. Les résultats obtenus de cette méthode sur les simulations sont présentés dans la section d'après.

6.1 Méthode de validation préliminaire

Lors de la modélisation, L-py produit une chaîne de caractères qui décrit la structure simulée. Cette chaîne de caractères peut être traduite au format MTG (Multiscale Tree Graph). Les UCs et les inflorescences sont placées à la même échelle. Les valeurs d'attributs du MTG sont données pour les paramètres des modules de la chaîne (paramètres des modules dans L-py). Nous avons donc un MTG simulé qui a la même structure que le MTG mesuré. Cela permet de comparer la simulation aux données observées, pour vérifier si la simulation reproduit bien aux niveaux quantitatif, qualitatif et temporel les phénomènes observés.

La validation porte sur différentes variables ciblées pour des raisons biologiques et agronomiques. La comparaison des résultats aux données empiriques se fait par cycle de production et par arbre. Les variables intéressantes sont :

- le nombre d'UCs produites au cours d'un cycle
- le nombre d'UCs filles issues des UCs mères terminales du cycle précédent
- la distribution du nombre d'UCs filles produites par axe au cours d'un cycle
- la distribution du nombre d'UCs filles par UC mère
- le pourcentage d'UCs terminales en position apicale à la fin de la période de croissance végétative

Le pourcentage s'obtient par le calcul suivant

$$\frac{\text{nombre d'UCs terminales en position apicale à la fin de la croissance végétative}}{\text{nombre d'UCs terminales à la fin de la croissance végétative}}$$

- le taux de floraison des UCs terminales

Le taux s’obtient par le calcul suivant

$$\frac{\text{nombre d'UCs terminales qui donnent au moins une inflo au cours du cycle de production}}{\text{nombre d'UCs terminales à la fin de la croissance végétative}}$$

- la distribution des dates de sortie des UC, exprimées à l’échelle du mois

Le procédé de validation est le suivant, pour un arbre et pour un cycle de production :

- a. récupération des valeurs des variables à partir du MTG empirique
- b. simulation de 1000 arbres, chacun mis au format MTG
- c. récupération des variables simulées pour chaque arbre
- d. comparaison des résultats simulés avec les variables empiriques

La récupération des valeurs empiriques, pour les variables citées précédemment (étape a), a été appliquée sur les 35 arbres du MTG empirique. Pour chaque, variable, les valeurs ont été sauvegardées dans un fichier. Nous nous sommes focalisés, cependant, sur les arbres de la variété Cogshall.

Les 1 000 simulations (étape b) ont été appliquées, pour l’instant, sur un arbre chargé (B12). Les simulations ont été sauvegardées dans un dossier. Le début et la fin des simulations sont calés sur les mêmes dates que celles de l’observation des arbres : juin 2003 à novembre 2005 (fin du cycle 5 avec les inflorescences). Le départ des simulations se fait en partant de la structure mesuré en 2003 d’arbre existant. Les inflorescences du cycle 3, poussant entre août et octobre 2003, ont été placées. N’ayant pas d’informations sur le nombre et la date d’apparition de ces inflorescences, il apparaît une seule inflorescence en position apicale en septembre 2003.

La récupération des variables simulées (étape c) a été effectuée en même temps que l’étape b.

La comparaison (étape d) entre les valeurs simulées et les valeurs empirique est possible, car le début des simulations est la même que pour les données empiriques (même date de départ et même nombre d’UCs au départ). Nous avons pour l’instant effectué des comparaisons préliminaires. Pour chaque variable citée précédemment, nous avons d’une part récupéré la valeur pour un arbre précis (en l’occurrence, ici l’arbre B12) et d’autre part nous avons une distribution de la variable simulée.

La comparaison préliminaires consiste à vérifier si la valeur empirique (généralement représenté en rouge dans les résultats) se trouve dans l’intervalle de 95% des valeurs prises par la simulation. Dans les résultats, nous plaçons également sur cet intervalle la médiane des valeurs prises (en vert), pour évaluer qualitativement une sous estimation ou une sur estimation de la simulation.

Pour les distributions des variables empiriques, nous affichons leurs distributions respectives (en rouge), puis nous ajoutons un boxplot des valeurs simulées sur chaque valeurs prises par la variable.

La validation a été réalisée sur les simulations utilisant les GLM et celles basées sur les modèles graphiques. Nous avons validé séparément les simulations basées sur les GLM complets et sur les GLM sélectionnés. Le but est de vérifier si l’amélioration de chaque

GLM complet en GLM sélectionné, améliore le modèle de croissance.

6.2 Premiers résultats

6.2.1 Le nombre d'UCs produites au cours d'un cycle

Au cycle 4

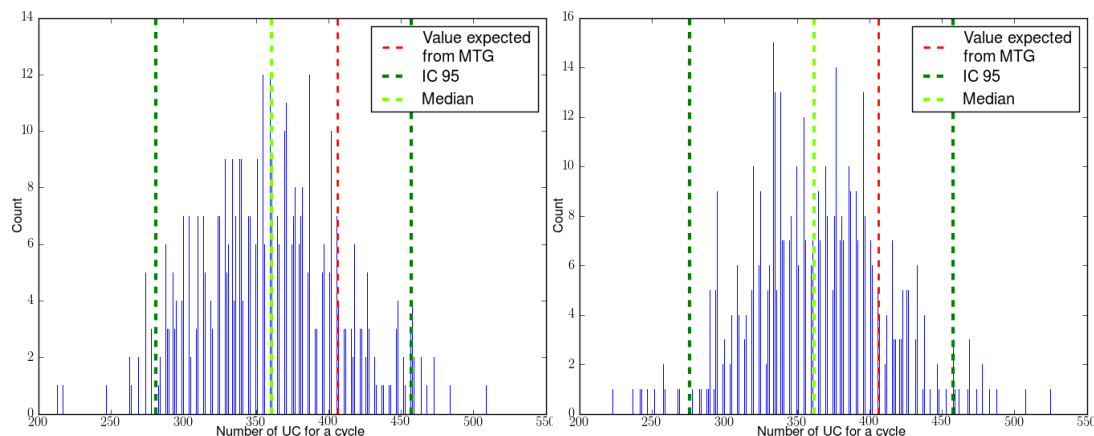


FIGURE 41 – Distribution du nombre d'UCs produites au cycle 4, à gauche à l'aide des GLM complets et à droite à l'aide des GML sélectionnés

Il n'y a pas de différences visibles pour le nombre d'UCs produites au cycle 4 entre les GLM sélectionnés et complets. Globalement, il y a une sous estimation de la populations à 50 UCs près. Pour les modèles graphiques, nous avons globalement une sur-estimation de la population.

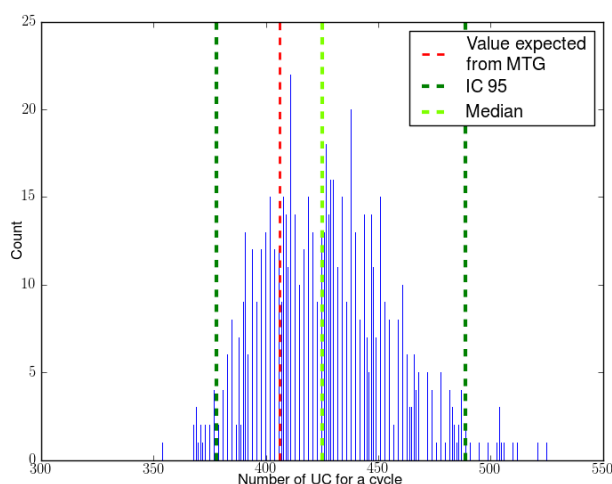


FIGURE 42 – Distribution du nombre d'UCs produites au cycle 4, à partir des modèles graphiques

Au cycle 5

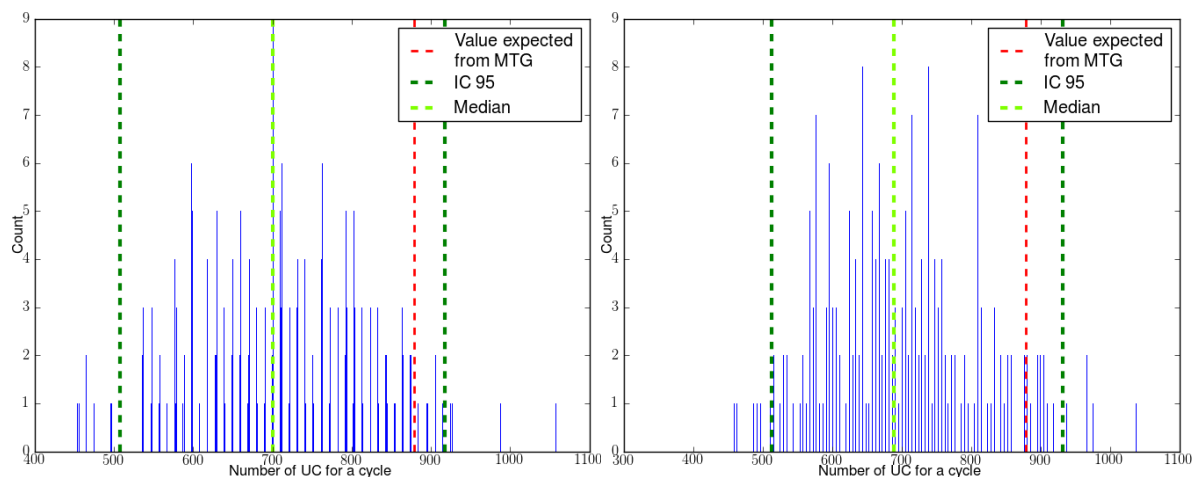


FIGURE 43 – Distribution du nombre d'UC produites au cycle 5, à gauche à l'aide des GLM complets et à droite à l'aide des GLM sélectionnés

Pour les GLMs, nous obtenons encore une sous-estimation de la population. Pour les modèles graphiques, il y a de grandes valeurs aberrantes. Ce modèle sur-estime la population globale des UCs. Cela peut être dû à l'utilisation de la régression de Poisson. Puisque les valeurs possibles de cette régression ne sont pas bornées. De plus, la fonction de lien inverse, utilisée pour la régression de Poisson, est l'exponentielle, elle fait donc fortement augmenter le nombre d'enfant d'une UC mère.

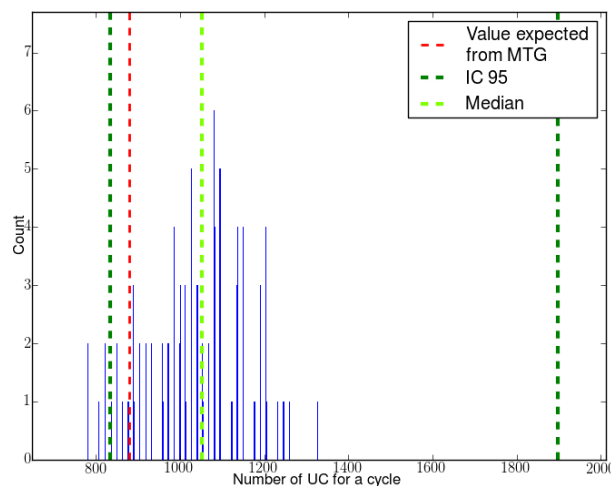


FIGURE 44 – Distribution du nombre d'UCs produites au cycle 5, à partir des modèles graphiques

6.2.2 La distribution du nombre d'UCs filles par UC mère

Au cycle 4

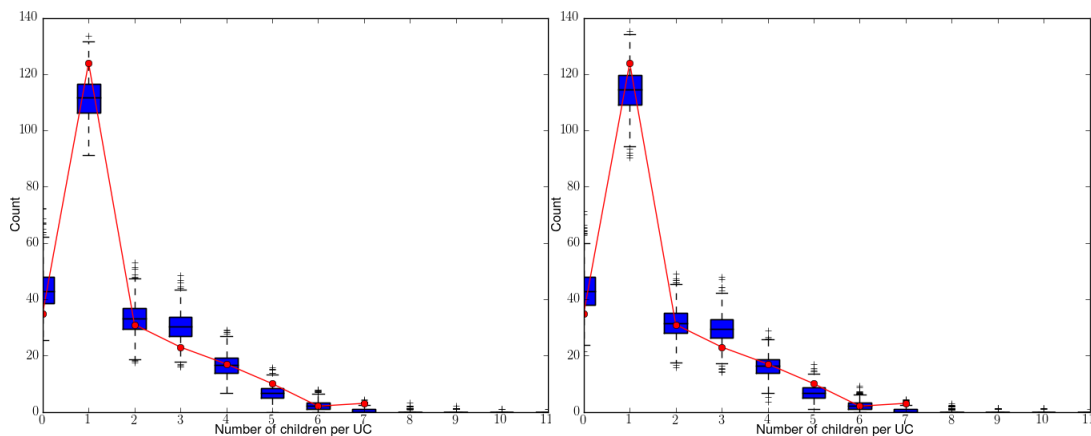


FIGURE 45 – Distribution du nombre d'UCs filles par UC mère au cycle 4, à gauche à l'aide des GLM complets et à droite à l'aide des GLM sélectionnés

Il n'y a pas de différences visibles pour le nombre d'UCs filles par UC mère entre les GLM sélectionnés et complets. Nous pouvons voir que les valeurs simulées sont proches des valeurs empiriques. Sur l'arbre B12, il y a maximum 7 UCs filles par UC mère, tandis que les 1 000 simulations vont jusqu'à 11 UCs filles. Pour les modèles graphiques, l'estimation du nombre d'enfants s'éloigne un peu plus des données observées. Le problème majeur est le nombre maximal d'UCs filles (26) qu'il n'est pas possible d'observer.

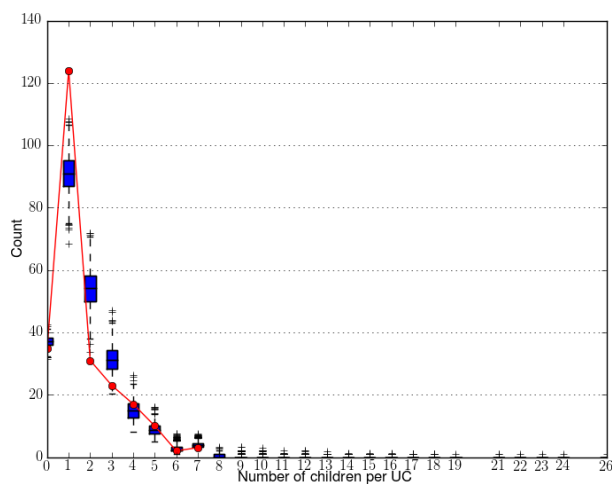


FIGURE 46 – Distribution du nombre d'UCs filles par UC mère au cycle 4, à partir des modèles graphiques

Au cycle 5

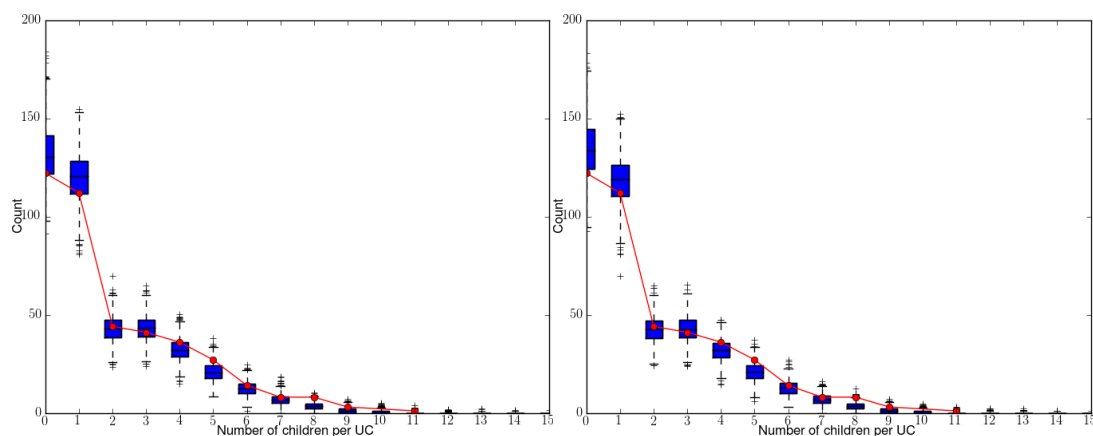


FIGURE 47 – Distribution du nombre d’UCs filles par UC mère au cycle 5, à gauche à l’aide des GLM complets et à droite à l’aide des GLM sélectionnés

Pour les GLM complets et sélectionnés, il n’y a toujours pas de différences visibles. Le nombre d’UCs filles simulé reste plus grand que le nombre maximal observé mais cela reste acceptable. En revanche, pour les modèles graphiques, l’ajustement semble de moins bonne qualité et nous avons des valeurs aberrantes : au maximum il y a 3 000 UCs fille pour une UC mère. Ce résultat a permis d’identifier des problèmes dans l’estimation des paramètres des modèles graphiques. Des corrections sont en cours pour améliorer cela.

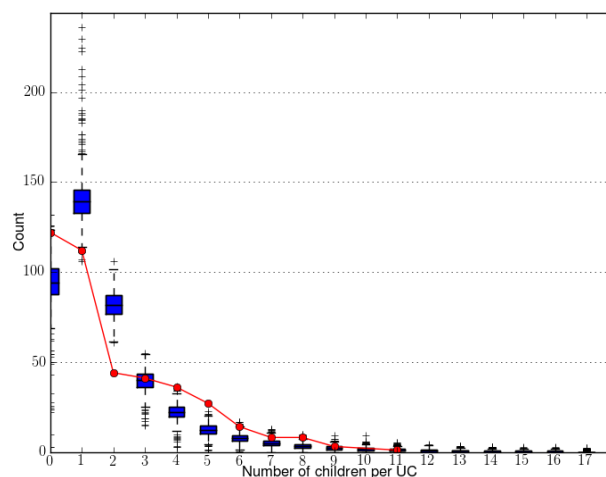


FIGURE 48 – Distribution du nombre d’UCs filles par UC mère au cycle 5, à partir des modèles graphiques

6.2.3 La distribution des dates d'apparition des UCs

Au cycle 4

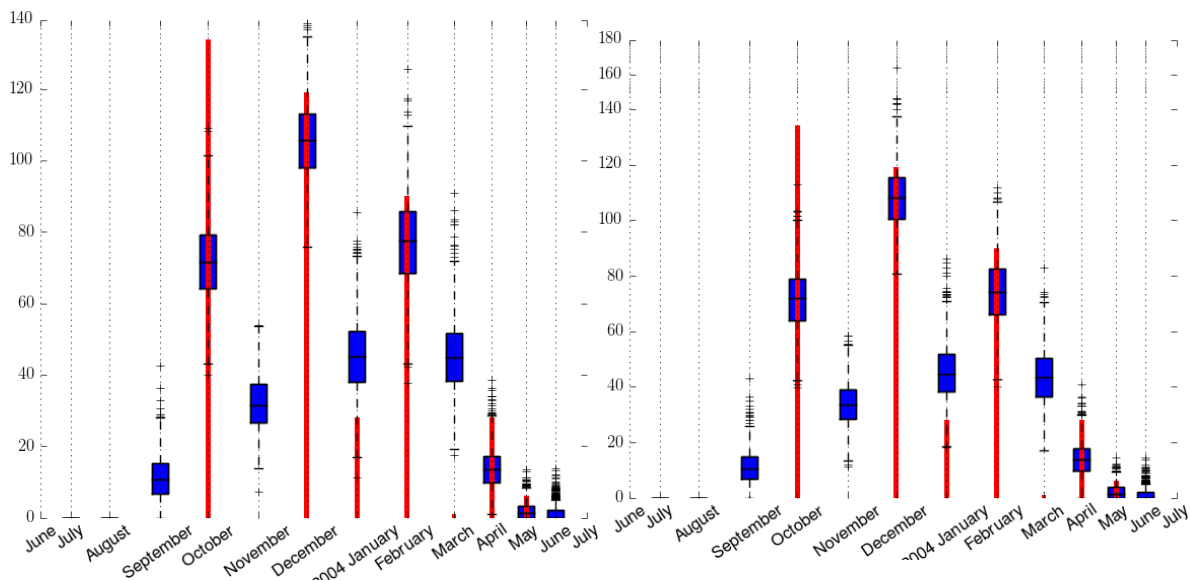


FIGURE 49 – Distribution des dates d'apparition des UCs au cycle 4, à gauche à l'aide des GLMs complets et à droite à l'aide des GLMs sélectionnés

Les histogrammes en rouge représentent la distribution empirique. Les boxplots en bleu indiquent la quantité d'UCs ayant débourrées à un mois donné.

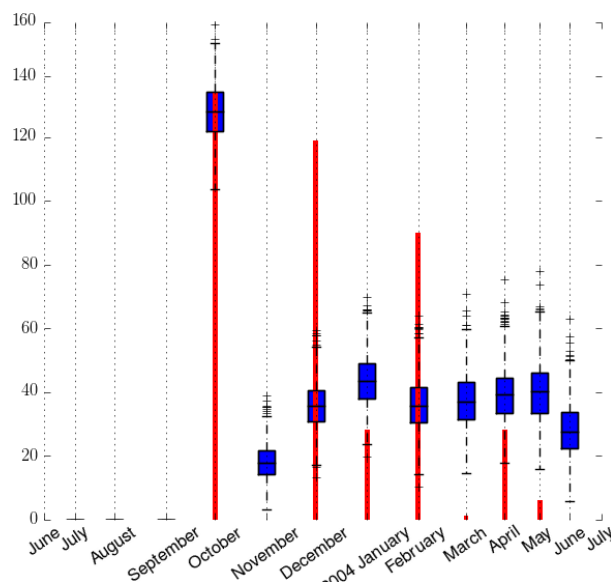


FIGURE 50 – Distribution des dates d'apparition des UCs au cycle 4, à partir des modèles graphiques

Les simulations basées sur les GLM montrent une mauvaise estimation des dates d'apparition des UCs. Nous avons utilisé une différence de temps entre la date d'apparition de l'UC mère et de ses UCs filles comme variable mais cela ne semble pas robuste. Une

alternative serait d'utiliser directement les dates d'apparition des UCs filles comme variable. Pour le modèle graphique, il y a aussi une mauvaise estimation de ces dates. La loi uniforme a été utilisée pour choisir un mois dans une période. Pour améliorer cela, il faudrait aussi utiliser la distribution empirique de ces dates sur les trois périodes Early, Intermediate et Late.

Au cycle 5

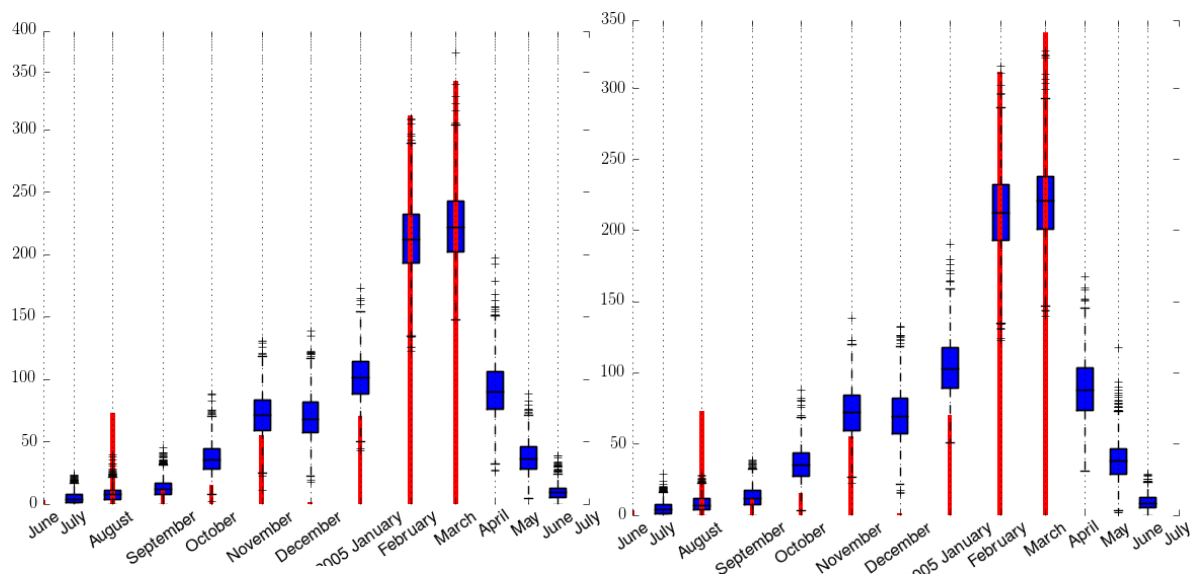


FIGURE 51 – Distribution des dates d'apparition des UCs au cycle 5, à gauche à l'aide des GLMs complets et à droite à l'aide des GLMs sélectionnés

Pour les simulations basées sur les GLM autant que sur les modèles graphiques, l'estimation des mois d'apparition est qualitativement de mauvaise qualité.

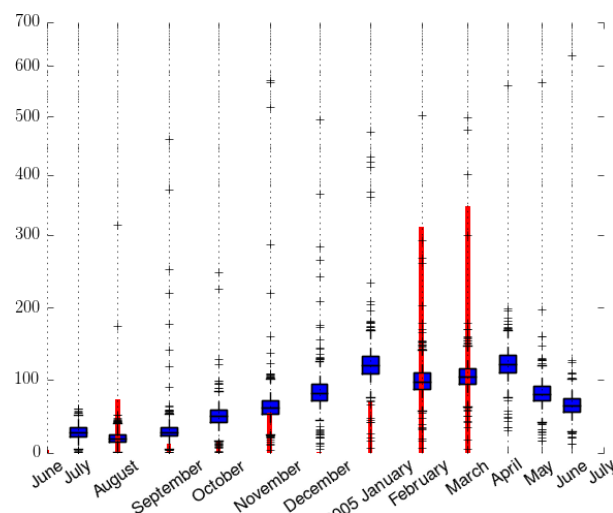


FIGURE 52 – Distribution des dates d'apparition des UCs au cycle 5, à partir des modèles graphiques

6.2.4 La distribution du taux d'UC terminale en position apicale

Au cycle 4

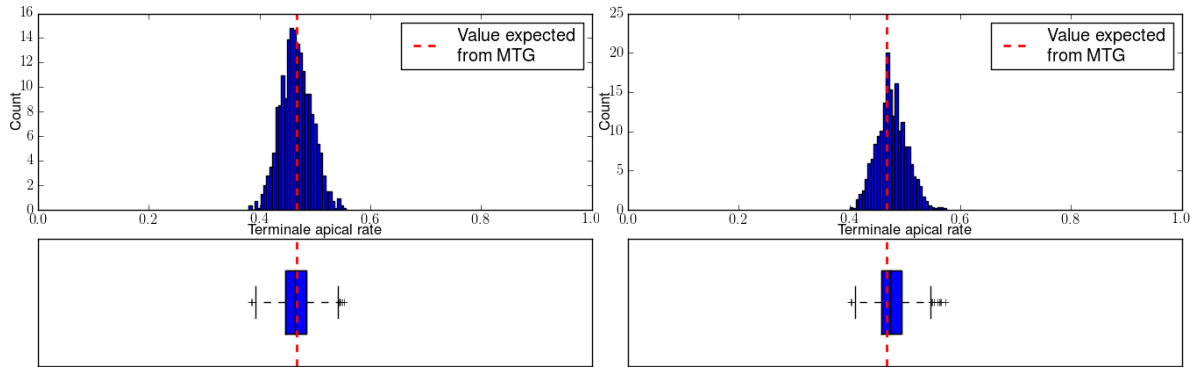


FIGURE 53 – Distribution du taux d'UCs terminales en position apicale au cycle 4, à gauche à l'aide des GLMs complets et à droite à l'aide des GLMs sélectionnés

Les simulations basées sur les GLM estiment correctement le taux d'UCs terminales en position apicale. L'estimation semble meilleure pour les GLMs complets que pour les GLM sélectionnés. Pour les modèles graphiques, il y a une surestimation des UCs terminales en position apicale.

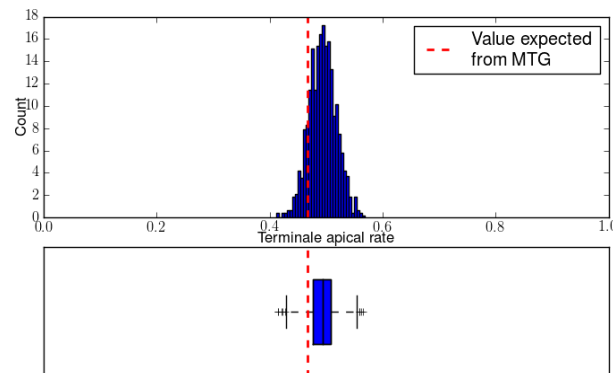


FIGURE 54 – Distribution du taux d'UC terminale en position apicale au cycle 4, à partir des modèles graphiques

Au cycle 5

Au cycle 5, il y a globalement une surestimation des UCs terminales en position apicale.

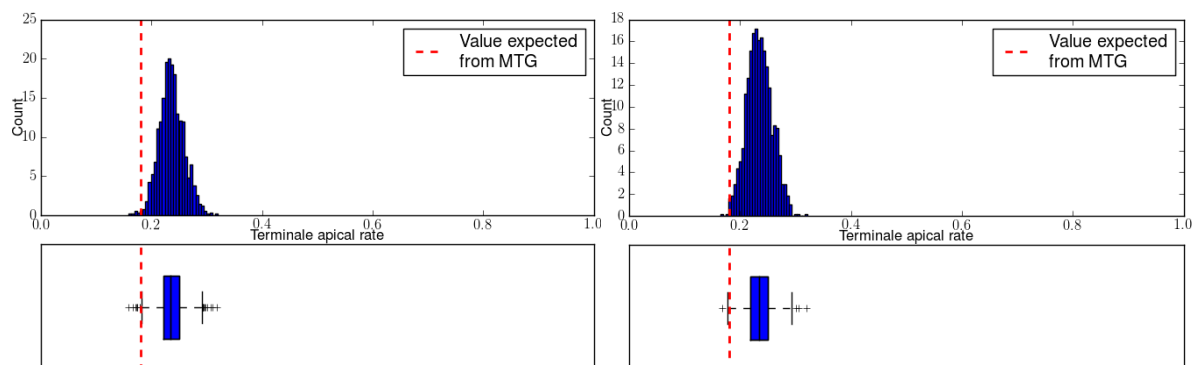


FIGURE 55 – Distribution du taux d’UC terminale en position apicale au cycle 5, à gauche à l’aide des GLMs complets et à droite à l’aide des GLMs sélectionnés

Puisque cette surestimation est inchangée selon les modèles choisis, il est possible que l’arbre B12 possède en moyenne moins d’UCs terminales apicale que les autres arbres Cogshall chargés sur lesquels les modèles ont été estimés. Pour vérifier cela, il faudrait appliquer la simulation sur les autres arbres Cogshall.

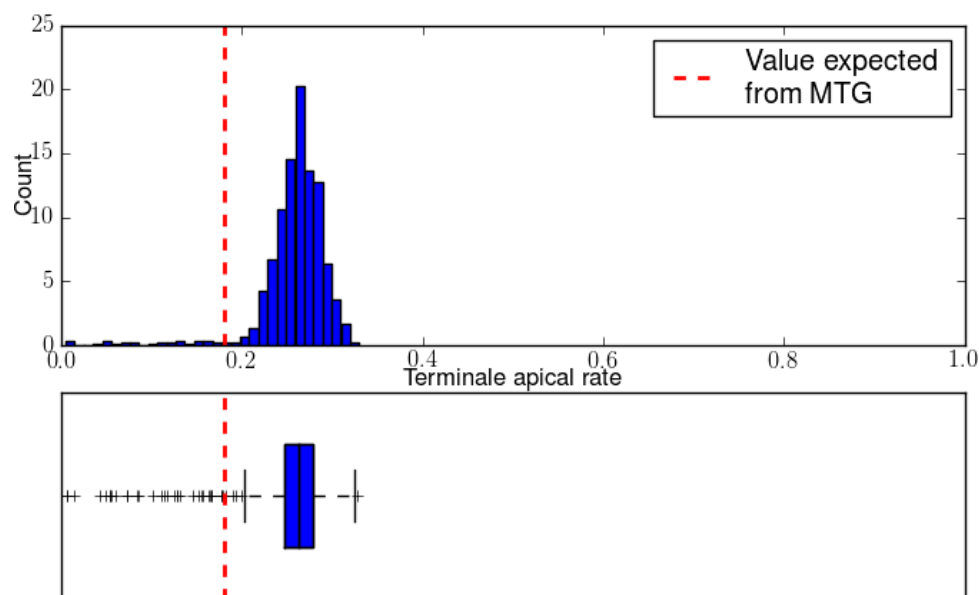


FIGURE 56 – Distribution du taux d’UC terminale en position apicale au cycle 5, à partir des modèles graphiques

6.2.5 La distribution du taux de floraison

Au cycle 4

Le taux de floraison est globalement sous-estimé.

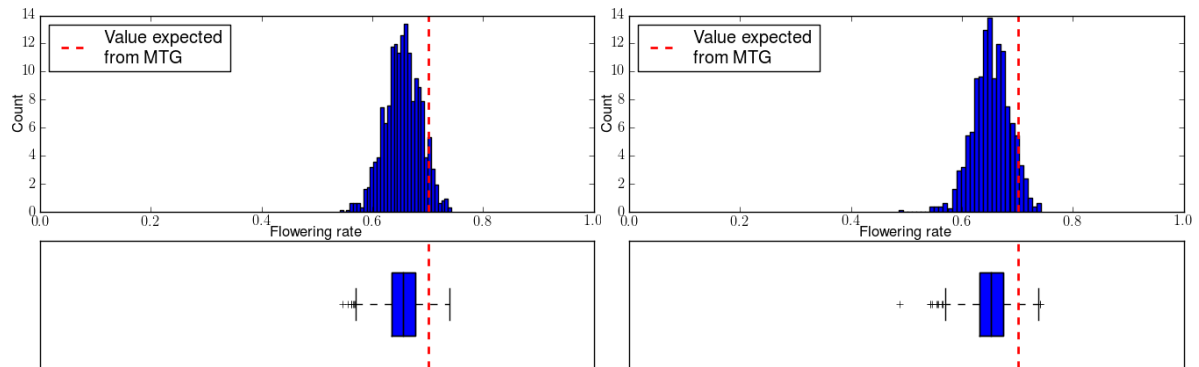


FIGURE 57 – Distribution du taux de floraison au cycle 4, à gauche à l'aide des GLMs complets et à droite à l'aide des GLMs sélectionnés

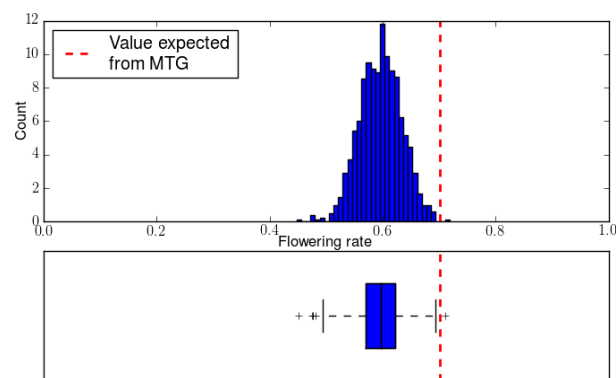


FIGURE 58 – Distribution du taux d'UC terminale en position apicale au cycle 4, à partir des modèles graphiques

Au cycle 5

Le taux de floraison reste sous-estimé pour les GLM, et surestimé pour les modèles graphiques. Ce changement pour les modèles graphiques peut s'expliquer par le fait qu'au cycle 5, nous avons eu beaucoup d'UCs produites au cours de ce cycle (jusqu'à 46 000 UCs). Les résultats sont donc biaisés par ce phénomène.

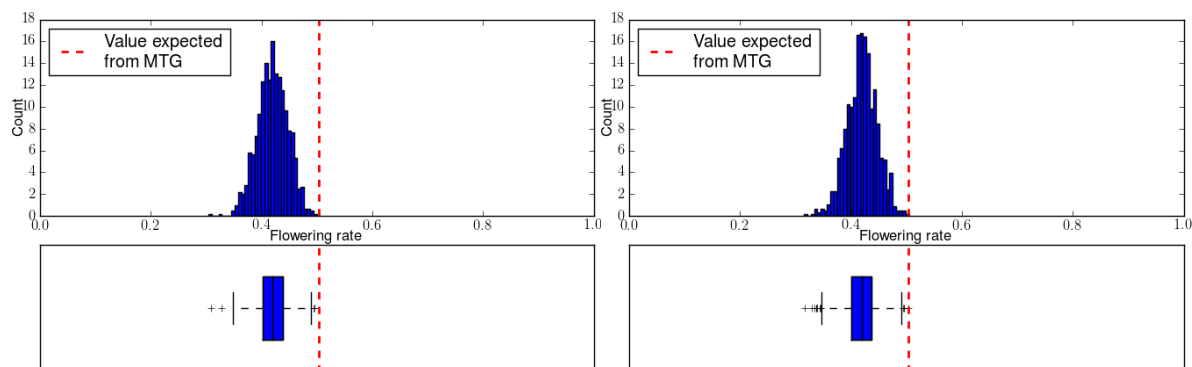


FIGURE 59 – Distribution du taux de floraison au cycle 5, à gauche à l'aide des GLMs complets et à droite à l'aide des GLMs sélectionnés

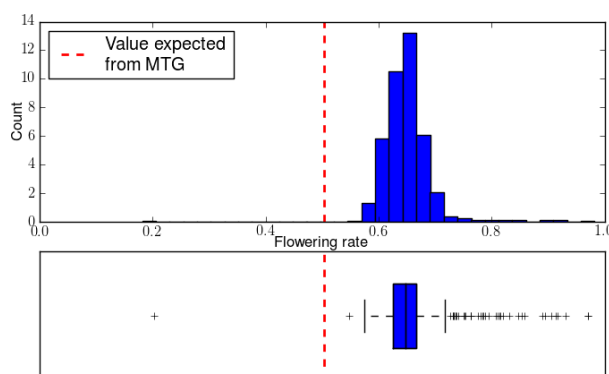


FIGURE 60 – Distribution du taux de floraison au cycle 5, à partir des modèles graphiques

Discussion et conclusion

Durant ce stage, nous avons mis en place un cadre conceptuel et informatique pour simuler l'architecture du manguier. Nous avons modélisé la croissance du manguier à l'aide du logiciel L-py en établissant une règle de croissance qui traduisait la relation entre une UC mère et ses UC filles. Cette règle de croissance nécessitait de connaître plusieurs informations : le nombre d'UC fille produite par l'UC mère, la date d'apparition des UC filles, le nombre d'inflorescences et la date de floraison. Ces informations ont été obtenues par deux types d'analyse statistique : les modèles linéaires généralisés et les modèles graphiques.

Les résultats préliminaires ont montré une bonne estimation du nombre d'UCs apparues dans un cycle, et du nombre d'UCs filles par UC mère. Mais il reste à améliorer les dates d'apparition des UCs, le nombre d'UCs terminales apicale et le nombre d'UCs terminales florifère. Plusieurs propositions sont faites pour les améliorations des modèles basés sur les GLM et sur les modèles graphiques.

Pour les modèles graphiques, nous pouvons mettre en place des distributions mélangées. Une amélioration de l'estimation du nombre d'UCs filles par UC mère, peut être obtenu en remplaçant, pour la régression de Poisson, la fonction de lien exponentielle par la fonction de lien identité en faisant attention à ne pas obtenir de valeur négative lors de la prédiction.

En ce qui concerne les GLMs, nous pouvons intégrer les interactions des facteurs aux GLMs. Cela donnera une information supplémentaire aux GLMs. Il reste aussi à vérifier si la quantité d'inflorescences est bien estimée dans le modèle de simulation.

Nous pourrions aussi améliorer la simulation de croissance en intégrant un phénomène observé : pour de jeune manguier qui ont une pousse régulière (car trop jeune pour faire des inflorescences), la date d'apparition des UCs filles ayant la même UCs grand-mère, est la même. Nous pouvons observer que le lien de parenté entre UCs s'étend aussi aux UCs grand-mère pour la temporalité. Ce phénomène peut persister à l'adolescence de l'arbre. Il faudrait donc prendre en compte ce synchronisme. Pour cela, dans L-py, les règles contextuelles peuvent aider à récupérer les informations de la structure. Les GLM seront modifiés en prenant un facteur supplémentaire.

Ce phénomène observé est important car il peut être la source d'un phénomène de groupement temporel et/ou spatial (patchiness) d'apparition des UCs. L'étude du patchiness permettrait de mieux comprendre comment le phénomène de groupement se met en place. Cette étude est essentielle pour connaître les actions qui permettraient de le contrer et améliorer les pratiques agronomiques. Le modèle proposé dans ce travail est un excellent support pour étudier dans les détails ce phénomène de patchiness.

Bibliographie

[Boudon, 2004] Boudon F. (2004). Représentation géométrique multi-échelles de l'architecture des plantes.

[Boudon, et al., 2012] Boudon F., Pradal C., Cokelaer T., Prusinkiewicz P., Godin C. (2012). L-Py : an L-system simulation framework for modeling plant architecture development based on a dynamic language. *Frontiers in plant science*, article 76.

[Dambreville et al., 2013a] Dambreville A., Lauri P-E., Trottier C., Guédon Y., Normand F. (2013). Deciphering structural and temporal interplays during the architectural development of mango trees. *Journal of Experimental Botany*, 64 : 2467-2480.

[Dambreville et al., 2013b] Dambreville A., Normand F., Lauri P-E. (2013). Plant growth co-ordination in natura : a unique temperature controlled law among vegetative and reproductive organs in mango. *Functional Plant Biology*, 40 : 280-291.

[Davenport, 2009] Davenport T.L. (2009). Reproductive physiology. In : Litz R.E.(ed). *The Mango Botany, production and uses*. 2nd edition. Wallingford, UK, CAB, 97-169.

[Gerbaud, 2012] Gerbaud P. (2012). Mango. *FruiTrop* 197, 19-54.

[Normand et al., 2009] Normand F., Pambo Bello A.K., Trottier C., Lauri P.E. (2009). Is axis position within tree architecture a determinant of axis morphology, branching, flowering and fruiting? An essay in mango. *Annals of Botany*, 103 (8) : 1325-1336.

[Normand et al., 2011] Normand F., Michels T., Lechaudel M., Joas J., Vincenot D., Hoarau I., Desmulier X., Barc G. (2011). Approche intégrée de la filière mangue à La Réunion. *Innovations agronomiques*, 17 : 67-81.

[Nunez-Elisea et Davenport, 1995] Nunez-Elisea R., Davenport T.L. (1995). Effect of leaf age, duration of cool temperature treatment, and photoperiod on bud dormancy release and floral initiation in mango. *Scientia Horticulturae* 62, 63-73.

[Pambo Bello, 2006] Pambo Bello A.K. (2006). Étude et modélisation des facteurs affectant les processus élémentaires de ramification et de croissance végétative chez le manguier cv. Cogshall à la Réunion. Rapport de master 2, Université Claude Bernard Lyon 1, 87p.

[Pradal et al., 2008] Pradal C., Dufour-Kowalski S., Boudon F., Fournier C., Godin C. (2008). OpenAlea : A visual programming and component-based software platform for plant modelling. *Functional Plant Biology*, 35(9-10) : 751-760.

[Prusinkiewicz et al., 1997] Prusinkiewicz, P., Hammel, M., Hanan, J., and Mech, R., editors (1997). Visual models of plant development, volume 3 of *Handbook of Formal Languages*. Springer-Verlag, Berlin.

[Vincenot et Normand, 2009] Vincenot D. (ed), Normand F. (ed) (2009). Guide de production intégrée de mangues à La Réunion. CIRAD et Chambre d'Agriculture de la Réunion, Sait-Pierre, Ile de la Réunion.

Annexes

A Symboles interprétés par la tortue

Nous donnons ici pour information un ensemble basique de symboles interprétés par la tortue qui servent à la compréhension de ce document.

- *Symboles qui font avancer et dessiner la tortue*

$F(s)$, $G(s)$ Avance d'un pas de longueur s et trace un segment à partir de la position d'origine jusqu'à la nouvelle position.

$f(s)$, $g(s)$ Avance d'un pas de longueur s sans tracer de segment.

$@O(r)$ Dessine une sphère de rayon r à la position courante.

- *Symboles qui contrôlent l'orientation de la tortue dans l'espace*

$+(\theta)$ Tourne à droite d'un angle θ autour de l'axe U .

$-(\theta)$ Tourne à gauche d'un angle θ autour de l'axe U .

$\&(\theta)$ Tourne vers le bas d'un angle θ autour de l'axe L .

$\^(\theta)$ Tourne vers le haut d'un angle θ autour de l'axe L .

$/(\theta)$ Roule vers la gauche d'un angle θ autour de l'axe H .

$\backslash(\theta)$ Roule vers la droite d'un angle θ autour de l'axe H .

$|$ Tourne de 180° autour de l'axe U . Ce qui est équivalent à $+(180)$ ou $-(180)$.

- *Symboles pour modéliser des structures ramifiées*

$[$ Pousse l'état courant de la tortue (position, orientation et attribut de traçage) dans une pile.

$]$ Sort de l'état de la pile et l'attribue à la tortue.

Aucune ligne n'est dessinée, même si en général la position et l'orientation de la tortue ont changé.

- *Symboles qui changent les attributs de traçage*

$_(\omega)$ Définit l'épaisseur de trait à ω , ou augmente la valeur courante d'épaisseur de trait de la valeur d'incrément par défaut si aucun paramètre n'est donné.

$!(\omega)$ Définit l'épaisseur de trait à ω , ou réduit la valeur courante d'épaisseur de trait de la valeur de décrétement par défaut si aucun paramètre n'est donné.

$;(n)$ Définit l'index de la palette de couleur à n , ou augmente la valeur de l'index courant par la valeur d'incrément par défaut si aucun paramètre n'est donné.

$.(n)$ Définit l'index de la palette de couleur à n , ou réduit la valeur de l'index courant par la valeur de décrétement par défaut si aucun paramètre n'est donné.